PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 9/42, 1/14

A1

(11) International Publication Number: WO 91/17244

(43) International Publication Date: 14 November 1991 (14.11.91)

(21) International Application Number: PCT/DK91/00124

(22) International Filing Date: 8 May 1991 (08.05.91)

(30) Priority data:
1158/90 9 May 1990 (09.05.90) DK

(71) Applicant *(for all designated States except US)*: NOVO NORDISK A/S [DK/DK]; Novo Allé, DK-2880 Bagsvaerd (DK).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : WÖLDIKE, Helle, Fabricius [DK/DK]; Stendyssevej 12, DK-3540 Lynge (DK). HAGEN, Fred [US/US]; 1315 Lexington Way East, Seattle, WA 98112 (US). HJORT, Carsten, Mailand [DK/DK]; Gåseageren 43, Svogerslev, DK-4000 Roskilde (DK). HASTRUP, Sven [DK/DK]; Frederiksborgvej 10, 3th, DK-2400 Copenhagen NV (DK).

(74) Common Representative: NOVO NORDISK A/S; Patent Department, Novo Allé, DK-2880 Bagsvaerd (DK).

(81) Designated States: AT (European patent), AU, BE (European patent), BR, CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, KR, LU (European patent), NL (European patent), NO, SE (European patent), SU, US.

Published
With international search report.

(54) Title: AN ENZYME CAPABLE OF DEGRADING CELLULOSE OR HEMICELLULOSE

(α)

```
1                                   10
Xaa Xaa Gln Cys Gly Gly Xaa Xaa Xaa Xaa Gly Xaa Xaa Xaa Cys Xaa

            20                                              30
Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Asn Xaa Xaa Tyr Xaa Gln Cys Xaa
            -
Xaa
 -
```

(57) Abstract

A cellulose- or hemicellulose-degrading enzyme which is derivable from a fungus other than *Trichoderma* or *Phanerochaete*, and which comprises a carbohydrate binding domain homologous to a terminal A region of *Trichoderma reesei* cellulases, which carbohydrate binding domain comprises amino acid sequence (α) or a subsequence thereof capable of effecting binding of the enzyme to an insoluble cellulosic or hemicellulosic substrate.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		

AN ENZYME CAPABLE OF DEGRADING CELLULOSE OR HEMICELLULOSE

FIELD OF INVENTION

The present invention relates to a cellulose- or hemicellulose-degrading enzyme, a DNA construct coding for the enzyme, a method of producing the enzyme, and an agent for degrading cellulose or hemicellulose comprising the enzyme.

BACKGROUND OF THE INVENTION

Enzymes which are able to degrade cellulose have previously been suggested for the conversion of biomass into liquid fuel, gas and feed protein. However, the production of fermentable sugars from biomass by means of cellulolytic enzymes is not yet able to compete economically with, for instance, the production of glucose from starch by means of α-amylase due to the inefficiency of the currently used cellulolytic enzymes. Cellulolytic enzymes may furthemore be used in the brewing industry for the degradation of β-glucans, in the baking industry for improving the properties of flour, in paper pulp processing for removing the non-crystalline parts of cellulose, thus increasing the proportion of crystalline cellulose in the pulp, and in animal feed for improving the digestibility of glucans. A further important use of cellulolytic enzymes is for textile treatment, e.g. for reducing the harshness of cotton-containing fabrics (cf., for instance, GB 1 368 599 or US 4,435,307), for soil removal and colour clarification of fabrics (cf., for instance, EP 220 016) or for providing a localized variation in colour to give the fabrics a "stone-washed" appearance (cf., for instance, EP 307 564).

The practical exploitation of cellulolytic enzymes has, to some extent, been set back by the nature of the known cellulase preparations which are often complex mixtures of a variety of single cellulase components, and which may have a rather low specific activity. It is difficult to optimise the production

of single components in multiple enzyme systems and thus to implement industrial cost-effective production of cellulolytic enzymes, and their actual use has been hampered by difficulties arising from the need to employ rather large quantities of the enzymes to achieve the desired effect.

The drawbacks of previously suggested cellulolytic enzymes may be remedied by using single-component enzymes selected for a high specific activity.

Single-component cellulolytic enzymes have been isolated from, e.g. Trichoderma reesei (cf. Teeri et al., Gene 51, 1987, pp. 43-52; P.M. Abuja, Biochem. Biophys. Res. Comm. 156, 1988, pp. 180-185; and P.J. Kraulis, Biochemistry 28, 1989, pp. 7241-7257). The T. reesei cellulases have been found to be composed of a terminal A region responsible for binding to cellulose, a B region linking the A region to the core of the enzyme, and a core containing the catalytically active domain. The A region of different T. reesei cellulases has been found to be highly conserved, and a strong homology has also been observed with a cellulase produced by Phanerochaete chrysosporium (Sims et al., Gene 74, 1988, pp. 411-422).

SUMMARY OF THE INVENTION

It has surprisingly been found that other fungi, which are not closely related to either Trichoderma reesei or Phanerochaete chrysosporium, are capable of producing enzymes which contain a region which is homologous to the A region of T. reesei cellulases.

Accordingly, the present invention relates to a cellulose- or hemicellulose-degrading enzyme which is derivable from a fungus other than Trichoderma or Phanerochaete, and which comprises a carbohydrate binding domain homologous to a terminal A region of Trichoderma reesei cellulases, which carbohydrate binding domain comprises the following amino acid sequence

1 10
Xaa Xaa Gln Cys Gly Gly Xaa Xaa Xaa Xaa Gly Xaa Xaa Xaa Cys Xaa

20 30
Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Asn Xaa Xaa Tyr Xaa Gln Cys Xaa

Xaa

or a subsequence thereof capable of effecting binding of the enzyme to an insoluble cellulosic or hemicellulosic substrate. "Xaa" is intended to indicate variations in the amino acid sequence of the carbohydrate binding domain of different enzymes. A hyphen is intended to indicate a "gap" in the amino acid sequence (compared to other, similar enzymes).

In the present context, the term "cellulose" is intended to include soluble and insoluble, amorphous and crystalline forms of cellulose. The term "hemicellulose" is intended to include glucans (apart from starch), mannans, xylans, arabinans or polyglucuronic or polygalacturonic acid. The term "carbohydrate binding domain" ("CBD") is intended to indicate an amino acid sequence capable of effecting binding of the enzyme to a carbohydrate substrate, in particular cellulose or hemicellulose as defined above. The term "homologous" is intended to indicate a high degree of identity in the sequence of amino acids constituting the carbohydrate binding domain of the present enzyme and the amino acids constituting the A region found in T. reesei cellulases ("A region" is the term used to denote the cellulose (i.e. carbohydrate) binding domain of T. reesei cellulases).

It is currently believed that cellulose- or hemicellulose-degrading enzymes which contain a sequence of amino acids which is identifiable as a carbohydrate binding domain (or "A region" based on its homology to the A region of T. reesei cellulases possess certain desirable characteristics as a result of the function of the carbohydrate binding domain in the enzyme molecule which is to mediate binding to solid substrates (including cellulose) and consequently to enhance the activity

of such enzymes towards such substrates. The identification and preparation of carbohydrate binding domain-containing enzymes from a variety of microorganisms is therefore of considerable interest.

Cellulose- or hemicellulose-degrading enzymes of the invention may conveniently be identified by screening genomic or cDNA libraries of different fungi with a probe comprising at least part of the DNA encoding the A region of T. reesei cellulases. Due to the intraspecies (i.e. different T. reesei cellulases) and interspecies homology observed for the carbohydrate binding domains of different cellulose- or hemicellulose-degrading enzymes, there is reason to believe that this screening method constitutes a convenient way of isolating enzymes of current interest.

DETAILED DISCLOSURE OF THE INVENTION

Carbohydrate binding domain (CBD) containing enzymes of the invention may, in particular, be derivable from strains of Humicola, e.g. Humicola insolens, Fusarium, e.g. Fusarium oxysporum, or Myceliopthora, e.g. Myceliopthora thermophile.

Some of the variations in the amino acid sequence shown above appear to be "conservative", i.e. certain amino acids are preferred in these positions among the various CBD-containing enzymes of the invention. Thus, in position 1 of the sequence shown above, the amino acid is preferentially Trp or Tyr. In position 2, the amino acid is preferentially Gly or Ala. In position 7, the amino acid is preferentially Gln, Ile or Asn. In position 8, the amino acid is preferentially Gly or Asn. In position 9, the amino acid is preferentially Trp, Phe or Tyr. In position 10, the amino acid is preferentially Ser, Asn, Thr or Gln. In position 12, the amino acid is preferentially Pro, Ala or Cys. In position 13, the amino acid is preferentially Thr, Arg or Lys. In position 14, the amino acid is preferentially Thr, Cys or Asn. In position 18, the amino acid

is preferentially Gly or Pro. In position 19, the amino acid (if present) is preferentially Ser, Thr, Phe, Leu or Ala. In position 20, the amino acid is preferentially Thr or Lys. In position 24, the amino acid is preferentially Gln or Ile. In position 26, the amino acid is preferentially Gln, Asp or Ala. In position 27, the amino acid is preferentially Trp, Phe or Tyr. In position 29, the amino acid is preferentially Ser, His or Tyr. In position 32, the amino acid is preferentially Leu, Ile, Gln, Val or Thr.

Examples of specific CBD-containing enzymes of the invention are those which comprise one of the following amino acid sequences

Trp Gly Gln Cys Gly Gly Gln Gly Trp Asn Gly Pro Thr Cys Cys Glu Ala Gly Thr Thr Cys Arg Gln Gln Asn Gln Trp Tyr Ser Gln Cys Leu;

Trp Gly Gln Cys Gly Gly Ile Gly Trp Asn Gly Pro Thr Thr Cys Val Ser Gly Ala Thr Cys Thr Lys Ile Asn Asp Trp Tyr His Gln Cys Leu;

Trp Gly Gln Cys Gly Gly Ile Gly Phe Asn Gly Pro Thr Cys Cys Gln Ser Gly Ser Thr Cys Val Lys Gln Asn Asp Trp Tyr Ser Gln Cys Leu;

Trp Gly Gln Cys Gly Gly Asn Gly Tyr Ser Gly Pro Thr Thr Cys Ala Glu Gly - Thr Cys Lys Lys Gln Asn Asp Trp Tyr Ser Gln Cys Thr Pro;

Trp Gly Gln Cys Gly Gly Gln Gly Trp Gln Gly Pro Thr Cys Cys Ser Gln Gly - Thr Cys Arg Ala Gln Asn Gln Trp Tyr Ser Gln Cys Leu Asn;

Trp Gly Gln Cys Gly Gly Gln Gly Tyr Ser Gly Cys Thr Asn Cys Glu Ala Gly Ser Thr Cys Arg Gln Gln Asn Ala Tyr Tyr Ser Gln Cys Ile;

Trp Gly Gln Cys Gly Gly Gln Gly Tyr Ser Gly Cys Arg Asn Cys Glu Ser Gly Ser Thr Cys Arg Ala Gln Asn Asp Trp Tyr Ser Gln Cys Leu;

Trp Ala Gln Cys Gly Gly Asn Gly Trp Ser Gly Cys Thr Thr Cys Val Ala Gly Ser Thr Cys Thr Lys Ile Asn Asp Trp Tyr His Gln Cys Leu;

Trp Gly Gln Cys Gly Gly Gln Asn Tyr Ser Gly Pro Thr Thr Cys Lys Ser Pro Phe Thr Cys Lys Lys Ile Asn Asp Phe Tyr Ser Gln Cys Gln; or

Trp Gly Gln Cys Gly Gly Asn Gly Trp Thr Gly Ala Thr Thr Cys Ala Ser Gly Leu Lys Cys Glu Lys Ile Asn Asp Trp Tyr Tyr Gln Cys Val

The cellulose- or hemicellulose-degrading enzyme of the invention may further comprise an amino acid sequence which defines a linking B region (to use the nomenclature established for T. reesei cellulases) adjoining the carbohydrate binding domain and connecting it to the catalytically active domain of the enzyme. The B region sequences established so far for enzymes of the invention indicate that such sequences are characterized by being predominantly hydrophilic and uncharged, and by being enriched in certain amino acids, in particular glycine and/or asparagine and/or proline and/or serine and/or threonine and/or glutamine. This characteristic structure of the B region imparts flexibility to the sequence, in particular in sequences containing short, repetitive units of primarily glycine and asparagine. Such repeats are not found in the B region sequences of T. reesei or P. chrysosporium which contain B regions of the serine/threonine type. The flexible structure is believed to facilitate the action of the catalytically active domain of the enzyme bound by the A region to the insoluble substrate, and therefore imparts advantageous properties to the enzyme of the invention.

Specific examples of B regions contained in enzymes of the invention have the following amino acid sequences

Ala Arg Thr Asn Val Gly Gly Gly Ser Thr Gly Gly Gly Asn Asn Gly Gly Gly Asn Asn Gly Gly Asn Pro Gly Gly Asn Pro Gly Gly Asn Pro Gly Gly Asn Pro Gly Gly Asn Pro Gly Gly Asn Pro Gly Gly Asn Pro Gly Gly Asn Cys Ser Pro Leu;

Pro Gly Gly Asn Asn Asn Asn Pro Pro Pro Ala Thr Thr Ser Gln Trp Thr Pro Pro Pro Ala Gln Thr Ser Ser Asn Pro Pro Pro Thr Gly Gly Gly Gly Gly Asn Thr Leu His Glu Lys;

Gly Gly Ser Asn Asn Gly Gly Gly Asn Asn Asn Gly Gly Gly Asn Asn Asn Gly Gly Gly Gly Asn Asn Asn Gly Gly Gly Asn Asn Asn Gly Gly Gly Asn Thr Gly Gly Gly Ser Ala Pro Leu;

Val Phe Thr Cys Ser Gly Asn Ser Gly Gly Gly Ser Asn Pro Ser Asn Pro Asn Pro Pro Thr Pro Thr Thr Phe Ile Thr Gln Val Pro Asn Pro Thr Pro Val Ser Pro Pro Thr Cys Thr Val Ala Lys;

Pro Ala Leu Trp Pro Asn Asn Asn Pro Gln Gln Gly Asn Pro Asn Gln Gly Gly Asn Asn Gly Gly Gly Asn Gln Gly Gly Gly Asn Gly Gly Cys Thr Val Pro Lys;

Pro Gly Ser Gln Val Thr Thr Ser Thr Thr Ser Ser Ser Ser Thr Thr Ser Arg Ala Thr Ser Thr Thr Ser Ala Gly Gly Val Thr Ser Ile Thr Thr Ser Pro Thr Arg Thr Val Thr Ile Pro Gly Gly Ala Ser Thr Thr Ala Ser Tyr Asn;

Glu Ser Gly Gly Gly Asn Thr Asn Pro Thr Asn Pro Thr Asn Pro Thr Asn Pro Thr Asn Pro Thr Asn Pro Trp Asn Pro Gly Asn Pro Thr Asn Pro Gly Asn Pro Gly Gly Gly Asn Gly Gly Asn Gly Gly Asn Cys Ser Pro Leu; or

Pro Ala Val Gln Ile Pro Ser Ser Ser Thr Ser Ser Pro Val Asn Gln Pro Thr Ser Thr Ser Thr Thr Ser Thr Ser Thr Thr Ser Ser Pro Pro Val Gln Pro Thr Thr Pro Ser Gly Cys Thr Ala Glu Arg

In another aspect, the present invention relates to a carbohydrate binding domain homologous to a terminal A region of Trichoderma reesei cellulases, which carbohydrate binding domain comprises the following amino acid sequence

1 10
Xaa Xaa Gln Cys Gly Gly Xaa Xaa Xaa Xaa Gly Xaa Xaa Xaa Cys Xaa

20 30
Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Asn Xaa Xaa Tyr Xaa Gln Cys Xaa

Xaa

or a subsequence thereof capable of effecting binding of a protein to an insoluble cellulosic or hemicellulosic substrate.

Examples of specific carbohydrate binding domains are those with the amino acid sequence indicated above.

In a further aspect, the present invention relates to a linking B region derived from a cellulose- or hemicellulose-degrading enzyme, said region comprising an amino acid sequence enriched in the amino acids glycine and/or asparagine and/or proline and/or serine and/or threonine and/or glutamine. As indicated above, these amino acids may often occur in short, repetitive units. Examples of specific B region sequences are those shown above.

The present invention provides a unique oppportunity to "shuffle" the various regions of different cellulose- or hemicellulose-degrading enzymes, thereby creating novel combinations of the CBD, B region and catalytically active domain resulting in novel activity profiles of this type of enzymes. Thus, the enzyme of the invention may be one which comprises an amino acid sequence defining a CBD, which amino acid sequence is derived from one naturally occurring cellulose- or hemicellulose-degrading enzyme, an amino acid sequence defining a linking B region, which amino acid sequence is derived from another naturally occurring cellulose- or

hemicellulose-degrading enzyme, as well as a catalytically active domain derived from the enzyme supplying either the CBD or the B region or from a third enzyme. In a particular embodiment, the catalytically active domain is derived from an enzyme which does not, in nature, comprise any CBD or B region. In this way, it is possible to construct enzymes with improved binding properties from enzymes which lack the CBD and B regions.

The enzyme of the invention is preferably a cellulase such as an endoglucanase (capable of hydrolysing amorphous regions of low crystallinity in cellulose fibres), a cellobiohydrolase (also known as an exoglucanase, capable of initiating degradation of cellulose from the non-reducing chain ends by removing cellobiose units) or a β-glucosidase.

In a still further aspect, the present invention relates to a DNA construct which comprises a DNA sequence encoding a cellulose- or hemicellulose-degrading enzyme as described above.

A DNA sequence encoding the present enzyme may, for instance, be isolated by establishing a cDNA or genomic library of a microorganism known to produce cellulose- or hemicellulose-degrading enzymes, such as a strain of Humicola, Fusarium or Mycelopthora, and screening for positive clones by conventional procedures such as by hybridization to oligonucleotide probes synthesized on the basis of the full or partial amino acid sequence of the enzyme or probes based on the partial or full DNA sequence of the A region from T. reesei cellulases, as indicated above, or by selecting for clones expressing the appropriate enzyme activity, or by selecting for clones producing a protein which is reactive with an antibody raised against a native cellulose- or hemicellulose-degrading enzyme.

Alternatively, the DNA sequence encoding the enzyme may be prepared synthetically by established standard methods, e.g.

the phosphoamidite method described by S.L. Beaucage and M.H. Caruthers, Tetrahedron Letters 22, 1981, pp. 1859-1869, or the method described by Matthes et al., The EMBO J. 3, 1984, pp. 801-805. According to the phosphoamidite method, oligonucleotides are synthesized, e.g. in an automatic DNA synthesizer, purified, annealed, ligated and cloned in appropriate vectors.

Finally, the DNA sequence may be of mixed genomic and synthetic, mixed synthetic and cDNA or mixed genomic and cDNA origin prepared by ligating fragments of synthetic, genomic or cDNA origin (as appropriate), the fragments corresponding to various parts of the entire DNA construct, in accordance with standard techniques. Thus, it may be envisaged that a DNA sequence encoding the CBD of the enzyme may be of genomic origin, while the DNA sequence encoding the B region of the enzyme may be of synthetic origin, or vice versa; the DNA sequence encoding the catalytically active domain of the enzyme may conveniently be of genomic or cDNA origin. The DNA construct may also be prepared by polymerase chain reaction using specific primers, for instance as described in US 4,683,202 or R.K. Saiki et al., Science 239, 1988, pp. 487-491.

The present invention also relates to an expression vector which carries an inserted DNA construct as described above. The expression vector may suitably comprise appropriate promotor, operator and terminator sequences permitting the enzyme to be expressed in a particular host organism, as well as an origin of replication enabling the vector to replicate in the host organism in question.

The resulting expression vector may then be transformed into a suitable host cell, such as a fungal cell, a preferred example of which is a species of Aspergillus, most preferably Aspergillus oryzae or Aspergillus niger. Fungal cells may be transformed by a process involving protoplast formation and transformation of the protoplasts followed by regeneration of

the cell wall in a manner known per se. The use of Aspergillus as a host microorganism is described in EP 238,023 (of Novo Industri A/S), the contents of which are hereby incorporated by reference.

Alternatively, the host organisms may be a bacterium, in particular strains of Streptomyces and Bacillus, and E. coli. The transformation of bacterial cells may be performed according to conventional methods, e.g. as described in Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, 1989.

The screening of appropriate DNA sequences and construction of vectors may also be carried out by standard procedures, cf. Sambrook et al., op. cit.

The invention further relates to a method of producing a cellulose- or hemicellulose-degrading enzyme as described above, wherein a cell transformed with the expression vector of the invention is cultured under conditions conducive to the production of the enzyme, and the enzyme is subsequently recovered from the culture. The medium used to culture the transformed host cells may be any conventional medium suitable for growing the host cells in question. The expressed enzyme may conveniently be secreted into the culture medium and may be recovered therefrom by well-known procedures including separating the cells from the medium by centrifugation or filtration, precipitating proteinaceous components of the medium by means of a salt such as ammonium sulphate, followed by chromatographic procedures such as ion exchange chromatography, affinity chromatography, or the like.

By employing recombinant DNA techniques as indicated above, techniques of fermentation and mutation or other techniques which are well known in the art, it is possible to provide cellulose- or hemicellulose-degrading enzymes of a high purity and in a high yield.

The present invention further relates to an agent for degrading cellulose or hemicellulose, the agent comprising a cellulose- or hemicellulose-degrading enzyme as described above. It is contemplated that, dependent on the specificity of the enzyme, it may be employed for one (or possibly more) of the applications mentioned above. In a particular embodiment, the agent may comprise a combination of two or more enzymes of the invention or a combination of one or more enzymes of the invention with one or more other enzymes with cellulose- or hemicellulose-degrading activity.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows the construction of plasmid p SX224;

Fig. 2 shows the construction of plasmid pHW485;

Fig. 3 shows the construction of plasmid pHW697 and pHW704;

Fig. 4 shows the construction of plasmid pHw768;

Fig. 5 is a restriction map of plasmid pSX320;

Fig. 6 shows the construction of plasmid pSX777

Fig. 7 shows the construction of plasmid pCaHj170;

Fig. 8 shows the construction of plasmid IM4;

Fig. 9 shows the SOE fusion of the ~43kD endoglucanase signal peptide and the N-terminal of Endo1;

Fig. 10 shows the construction of plasmid pCaHj180;

Fig. 11 shows the DNA sequence and derived amino acid sequence of F.oxysporum C-family cellobiohydrolase;

Fig. 12 shows the DNA sequence and derived amino acid sequence of F.oxysporum F-family cellulase;

Fig. 13 shows the DNA sequence and derived amino acid sequence of F.oxysporum C-family endoglucanase;

Fig. 14.A-E whows the DNA sequence and derived amino acid sequence of H.insolens endoglucanase 1(EG1); and

Fig. 15A-D shows the DNA sequence and derived amino acid sequence of a fusion of the B.lautus (NCIMB 40250) Endo 1 catalytic domain and the CBD and B region of H.insolens ~43kD endoglucanase.

The invention is further illustrated in the following examples which are not in any way intended to limit the scope of the invention as claimed.

Example 1

Isolation of A region-containing clones from H. insolens

From H. insolens strain DSM 1800 (described in, e.g. WO 89/09259) grown on cellulose, mRNA was prepared according to the method described by Koplan et al., Biochem. J. 183 (1979) 181-184. A cDNA library containing 20,000 clones was obtained substantially by the method of Okayama and Berg, Methods in Enzymology 154, 1987, pp. 3-28.

The cDNA library was screened as described by Gergen et al., Nucl. Acids Res. 7(8), 1979, pp. 2115-2136, with oligonucleotide probes in the antisense configuration, designed according to the published sequences of the N-terminal part of the A-region of the four T. reesei cellulase genes (Penttilä et al., Gene 45 (1986), 253-63; Saloheimo et al., Gene 63, (1988), 11-21; Shoemaker et al., Biotechnology, October 1983, 691-696; Teeri et al., Gene 51 (1987) 43-52. The probe sequences were as follows:

NOR-804 (~ EG 1)	5'-CTT GCA CCC GCT GTA CCC AAT GCC ACC GCA CTG CCC CCA-3'
NOR-805 (~CBH 1)	5'-CGT GGG GCC GCT GTA GCC AAT ACC GCC GCA CTG GCC GTA-3'
NOR-807 (~CBH 2)	5'-AGT CGG ACC CGA CCA ATT CTG GCC ACC ACA TTG GCC CCA-3'
NOR-808 (~EG 3)	5'-CGT AGG TCC GCT CCA ACC AAT ACC TCC ACA CTG GCC CCA-3'

Screening yielded a large number of candidates hybridising well to the A-region probes. Restriction mapping reduced the number of interesting clones to 17, of which 8 have so far been

sequenced (as described by Haltiner et al., Nucl. Acids Res. 13, 1985, pp. 1015-1025) sufficiently to confirm the presence of a terminal CBD as well as a B-region.

The deduced amino acid sequences obtained for the CBDs were as follows

A-1: Trp Gly Gln Cys Gly Gly Gln Gly Trp Asn Gly Pro Thr Cys Cys Glu Ala Gly Thr Thr Cys Arg Gln Gln Asn Gln Trp Tyr Ser Gln Cys Leu;

A-5: Trp Gly Gln Cys Gly Gly Ile Gly Trp Asn Gly Pro Thr Thr Cys Val Ser Gly Ala Thr Cys Thr Lys Ile Asn Asp Trp Tyr His Gln Cys Leu;

CBH-2: Trp Gly Gln Cys Gly Gly Ile Gly Phe Asn Gly Pro Thr Cys Cys Gln Ser Gly Ser Thr Cys Val Lys Gln Asn Asp Trp Tyr Ser Gln Cys Leu;

A-8: Trp Gly Gln Cys Gly Gly Asn Gly Tyr Ser Gly Pro Thr Thr Cys Ala Glu Gly - Thr Cys Lys Lys Gln Asn Asp Trp Tyr Ser Gln Cys Thr Pro;

A-9: Trp Gly Gln Cys Gly Gly Gln Gly Trp Gln Gly Pro Thr Cys Cys Ser Gln Gly - Thr Cys Arg Ala Gln Asn Gln Trp Tyr Ser Gln Cys Leu Asn;

A-11: Trp Gly Gln Cys Gly Gly Gln Gly Tyr Ser Gly Cys Thr Asn Cys Glu Ala Gly Ser Thr Cys Arg Gln Gln Asn Ala Tyr Tyr Ser Gln Cys Ile;

A-19: Trp Gly Gln Cys Gly Gly Gln Gly Tyr Ser Gly Cys Arg Asn Cys Glu Ser Gly Ser Thr Cys Arg Ala Gln Asn Asp Trp Tyr Ser Gln Cys Leu; and

43 kD: Trp Ala Gln Cys Gly Gly Asn Gly Trp Ser Gly Cys Thr Thr Cys Val Ala Gly Ser Thr Cys Thr Lys Ile Asn Asp Trp Tyr His Gln Cys Leu

The deduced amino acid sequences obtained for the B region were as follows

A1: Ala Arg Thr Asn Val Gly Gly Gly Ser Thr Gly Gly Gly Asn Asn Gly Gly Gly Asn Asn Gly Gly Asn Pro Gly Gly Asn Pro Gly Gly Asn Pro Gly Gly Asn Pro Gly Gly Asn Pro Gly Gly Asn Pro Gly Gly Asn Cys Ser Pro Leu;

A5: Pro Gly Gly Asn Asn Asn Asn Pro Pro Pro Ala Thr Thr Ser Gln Trp Thr Pro Pro Pro Ala Gln Thr Ser Ser Asn Pro Pro Pro Thr Gly Gly Gly Gly Gly Asn Thr Leu His Glu Lys;

A8: Gly Gly Ser Asn Asn Gly Gly Gly Asn Asn Asn Gly Gly Gly Asn Asn Asn Gly Gly Gly Asn Asn Asn Gly Gly Gly Asn Asn Asn Gly Gly Gly Asn Thr Gly Gly Gly Ser Ala Pro Leu;

A11: Val Phe Thr Cys Ser Gly Asn Ser Gly Gly Gly Ser Asn Pro Ser Asn Pro Asn Pro Pro Thr Pro Thr Thr Phe Ile Thr Gln Val Pro Asn Pro Thr Pro Val Ser Pro Pro Thr Cys Thr Val Ala Lys;

A19: Pro Ala Leu Trp Pro Asn Asn Asn Pro Gln Gln Gly Asn Pro Asn Gln Gly Gly Asn Asn Gly Gly Gly Asn Gln Gly Gly Gly Asn Gly Gly Cys Thr Val Pro Lys;

CBH2: Pro Gly Ser Gln Val Thr Thr Ser Thr Thr Ser Ser Ser Ser Thr Thr Ser Arg Ala Thr Ser Thr Thr Ser Ala Gly Gly Val Thr Ser Ile Thr Thr Ser Pro Thr Arg Thr Val Thr Ile Pro Gly Gly Ala Ser Thr Thr Ala Ser Tyr Asn;

A9: Glu Ser Gly Gly Gly Asn Thr Asn Pro Thr Asn Pro Thr Asn Pro Thr Asn Pro Thr Asn Pro Thr Asn Pro Trp Asn Pro Gly Asn Pro Thr Asn Pro Gly Asn Pro Gly Gly Gly Asn Gly Gly Asn Gly Gly Asn Cys Ser Pro Leu; or

Pro Ala Val Gln Ile Pro Ser Ser Ser Thr Ser Ser Pro Val Asn Gln Pro Thr Ser Thr Ser Thr Thr Ser Thr Ser Thr Thr Ser Ser Pro Pro Val Gln Pro Thr Thr Pro Ser Gly Cys Thr Ala Glu Arg

Example 2

Expression in A. oryzae of a CBH 2-type cellulase from H. insolens

The complete sequence of one of the CBD clones shows a striking similarity to a cellobiohydrolase (CBH 2) from T. reesei.

The construction of the expression vector pSX224 carrying the H. insolens CBH 2 gene for expression in and secretion from A. oryzae is outlined in Fig. 1. The vector p777 containing the pUC 19 replicon and the regulatory regions of the TAKA amylase promoter from A. oryzae and glucoamylase terminator from A. niger is described in EP 238 023. pSX 217 is composed of the cloning vector pcDV1-pL1 (cf. Okayama and Berg, op. cit.) carrying the H. insolens CBH 2 gene on a 1.8 kb fragment. The CBH 2 gene contains three restriction sites used in the construction: A BalI site at the initiating methionine codon in the signal sequence, a BstBI site 620 bp downstream from the BalI site and an AvaII site 860 bp downstream from the BstBI site. The AvaII site is located in the non-translated C-terminal part of the gene upstream of the poly A region, which is not wanted in the final construction. Nor is the poly G region upstream of the gene in the cloning vector. This region is excised and replaced by an oligonucleotide linker which places the translational start codon close to the BamHI site at the end of the TAKA promoter.

The expression vector pSX 224 was transformed into A. oryzae IFO 4177 using the amdS gene from A. nidulans as the selective marker as described in EP 238 023. Transformants were grown in YPD medium (Sherman et al., Methods in Yeast Genetics, Cold

Spring Harbor Laboratory, 1981) for 3-4 days and analysed for new protein species in the supernatant by sodium dodecyl sulphate polyacrylamide gel electrophoresis. The CBH 2 from H. insolens formed a band with an apparent Mw of 65 kD indicating a substantial glycosylation of the protein chain, which is calculated to have a Mw of 51 kD on the basis of the amino acid composition. The intact enzyme binds well to cellulose, while enzymatic degradation products of 55 kD and 40 kD do not bind, indicating removal of the A-region and possibly the B-region. The enzyme has some activity towards filter paper, giving rise to release of glucose. As expected, it has very limited endoglucanase activity as measured on soluble cellulose in the form of carboxy methyl cellulose.

Example 3

Isolation of Fusarium oxysporum genomic DNA

A freeze-dried culture of Fusarium oxysporum was reconstituted with phosphate buffer, spotted 5 times on each of 5 FOX medium plates (6% yeast extract, 1.5% K_2HPO_4, 0.75% $MgSO_4$ $7H_2O$, 22.5% glucose, 1.5% agar, pH 5.6) and incubated at 37°C. After 6 days of incubation the colonies were scraped from the plates into 15 ml of 0.001% Tween-80 which resulted in a thick and cloudy suspension.

Four 1-liter flasks, each containing 300 ml of liquid FOX medium, were inoculated with 2 ml of the spore suspension and were incubated at 30°C and 240 rpm. On the 4th day of incubation, the cultures were filtered through 4 layers of sterile gauze and washed with sterile water. The mycelia were dried on Whatman filter paper, frozen in liquid nitrogen, ground into a fine powder in a cold mortar and added to 75 ml of fresh lysis buffer (10 mM Tris-Cl 7.4, 1% SDS, 50 mM EDTA, 100 µl DEPC). The thoroughly mixed suspension was incubated in a 65°C waterbath for 1 hour and then spun for 10 minutes at 4000 rpm and 5°C in a bench-top centrifuge. The supernatant

was decanted and EtOH precipitated. After 1 hour on ice the solution was spun at 19,000 rpm for 20 minutes. The supernatant was decanted and isopropanol precipitated. Following centrifugation at 10,000 rpm for 10 minutes, the supernatant was decanted and the pellets allowed to dry.

One milliliter of TER solution (10 mM Tris-HCl, pH 7.4, 1 mM EDTA, 100 µg RNAse A) was added to each tube, and the tubes were stored at 4°C for two days. The tubes were pooled and placed in a 65°C waterbath for 30 minutes to suspend non-dissolved DNA. The solution was extracted twice with phenol/$CHCl_3$/isoamyl alcohol, twice with $CHCl_3$/isoamyl alcohol and then ethanol precipitated. The pellet was allowed to settle and the EtOH was removed. 70% EtOH was added and the DNA stored overnight at -20°C. After decanting and drying, 1 ml of TER was added and the DNA was dissolved by incubating the tubes at 65°C for 1 hour. The preparation yielded 1.5 mg of genomic DNA.

Amplification, cloning and sequencing of DNA amplified with degenerate primers

To amplify DNA from C-family (according to the nomenclature of Henrissat et al. Gene 81 (1), 1989, pp. 83-96) cellulases using PCR (cf. US 4,683,195 and US 4,683,202) each "sense" oligonucleotide was used in combination with each "antisense" oligonucleotide. Thus, the following primer pair was used:

Primer 1	Primer 2
ZC3220	ZC3221

ZC3220: GCC AAC TAC GGT ACC GG(A/C/G/T) TA(C/T) TG(C/T) GA(C/T) (A/G/T)(C/G)(A/G/C/T) CA(G/A) TG

ZC3221: GCG TTG GCC TCT AGA AT(G/A) TCC AT(C/T) TC(A/G/C/T) (C/G/T)(A/T)(G/A) CA(G/A) CA

In the PCR reaction, 1 µg of Fusarium oxysporum genomic DNA was used as the template. Ten times PCR buffer is 100mM Tris-HCl pH 8.3, 500 mM KCl, 15 mM MgCl, 0.1% gelatin (Perkin-Elmer Cetus). The reactions contained the following ingredients:

dH2O	35.75 µl	
10X PCR buffer	5 µl	
template DNA	5 µl	
primer 1	2 µl	(40pmol)
primer 2	2 µl	(40 pmol)
Taq polymerase	0.25 µl	(1.25 U)
total	50 µl	

The PCR reactions were performed for 40 cycles under the following conditions:

94°C	1.5 min
45°	2.0 min
72°	2.0 min

Five microliters of each reaction was analyzed by agarose gel electrophoresis. The sizes of the DNA fragments were estimated from DNA molecular weight markers. The reacton primed with ZC3220 and ZC3221, produced two DNA fragments of appropriate size to be candidates for fragments of C-family cellulases. The agarose sections containing these two fragments were excised, and the DNA was electroeluted and digested with the restriction enzymes KpnI and ZbaI. The fragments were ligated into the vector pUC18 which had been cut with the same two restriction enzymes. The ligations were transformed into E. coli and mini-prep DNA was prepared from the resulting colonies. The DNA sequences of these inserts were determined and revealed that two new C-family cellulases had been identified, one a new cellobiohydrolase and the other a new endoglucanase.

The PCR cloning strategy described above for the C-family cellulases was applied using other primers which encoded conserved cellulase sequences within the known F-family

cellulases (cf. Henrissat et al., op. cit.) The following primer pair was used for amplification of Fusarium genomic DNA.

	Primer 1	Primer 2
	ZC3226	ZC3227

ZC3226: TCC TGA CGC CAA GCT TT(A/G/T) (C/T)(A/T)(A/T) (A/C/T)AA (C/T)GA (C/T)TA (C/T)AA

ZC3227: CAC CGG CAC CAT CGA T(G/A/)T C(A/C/G/T)A (G/A)(C/T)T C(A/G/C/T)G T(A/G/T)A T

The PCR reactions were performed for 40 cycles as follows:

94°C	1.5 min
50°C	2.0 min
72°C	2.0 min

The 180 bp band was eluted from an agarose gel fragment, digested with the restriction enzymes Hind III and Cla I and ligated into pUC19 which had been digested with Hind III and AccI. The ligated DNA was transformed into E. coli and mini-prep DNA was prepared from colony isolates. The DNA sequence of the cloned DNA was determined. This fragment encoded sequences corresponding to a new member of the F-family cellulases.

Construction of a Fusarium oxysporum cDNA library

Fusarium oxysporum was grown by fermentation and samples were withdrawn at various times for RNA extraction and cellulase activity analysis. The activity analysis included an assay for total cellulase activity as well as one for colour clarification. Fusarium oxysporum samples demonstrating maximal colour clarification were extracted for total RNA from which poly(A)+RNA was isolated.

To construct a Fusarium oxysporum cDNA library, first-strand cDNA was synthesized in two reactions, one with and the other without radiolabelled dATP. A 2.5X reaction mixture was prepared at room temperature by mixing the following reagents in the following order: 10 µl of 5X reverse transcriptase buffer (Gibco-BRL, Gaithersburg, Maryland) 2.5 µl 200 mM dithiothreitol (made fresh or from a stock solution stored at -70°C), and 2.5 µl of a mixture containing 10 mM of each deoxynucleotide triphosphate, (dATP, dGTP, dTTP and 5-methyl dCTP, obtained from Pharmacia LKB Biotechnology, Alameda, CA). The reaction mixture was divided into each of two tubes of 7.5 µl. 1.3 µl of 10 µCi/µl ^{32}P α-dATP (Amersham, Arlington Heights, IL) was added to one tube and 1.3 µl of water to the other. Seven microliters of each mixture was transferred to final reaction tubes. In a separate tube, 5 µg of Fusarium oxysporum poly $(A)^+$ RNA in 14 µl of 5 mM Tris-HCl pH 7.4, 50 µM EDTA was mixed with 2 µl of 1 µg/µl first strand primer (ZC2938 $GACAGAGCACAGAATTCACTAGTGAGCTCT_{15}$). The RNA-primer mixture was heated at 65°C for 4 minutes, chilled in ice water, and centrifuged briefly in a microfuge. Eight microliters of the RNA-primer mixture was added to the final reaction tubes. Five microliters of 200 U/µl Superscript™ reverse transcriptase (Gibco-BRL) was added to each tube. After gentle agitation, the tubes were incubated at 45°C for 30 minutes. Eighty microliters of 10 mM Tris-HCl pH 7.4, 1 mM EDTA was added to each tube, the samples were vortexed, and briefly centrifuged. Three microliters was removed from each tube to determine counts incorporated by TCA precipitation and the total counts in the reaction. A 2 µl sample from each tube was analyzed by gel electrophoresis. The remainder of each sample was ethanol precipitated in the presence of oyster glycogen. The nucleic acids were pelleted by centrifugation, and the pellets were washed with 80% ethanol. Following the ethanol wash, the samples were air dried for 10 minutes. The first strand synthesis yielded 1.6 µg of Fusarium oxysporum cDNA, a 33% conversion of poly(A)+RNA into DNA.

Second strand cDNA synthesis was performed on the RNA-DNA hybrid from the first strand reactions under conditions which encouraged first strand priming of second strand synthesis resulting in hairpin DNA. The first strand products from each of the two first strand reactions were resuspended in 71 μl of water. The following reagents were added, at room temperature, to the reaction tubes: 20 μl of 5X second strand buffer (100 mM Tris pH 7.4, 450 mM KCl, 23 mM $MgCl_2$, and 50 mM $(NH_4)_2(SO_4)$, 3 μl of 5 mM β-NAD, and μl of a deoxynucleotide triphosphate mixture with each at 10 mM. One microliter of α-^{32}P dATP was added to the reaction mixture which received unlabeled dATP for the first strand synthesis while the tube which received labeled dATP for first strand synthesis received 1 μl of water. Each tube then received 0.6 μl of 7 U/μl E. coli DNA ligase (Boehringer-Mannheim, Indianapolis, IN), 3.1 μl of 8 U/μl E. coli DNA polymerase I (Amersham), and 1 μl 2 U/μl of RNase H (Gibco-BRL). The reactions were incubated at 16°C for 2 hours. After incubation, 2μl from each reaction was used to determine TCA precipitable counts and total counts in the reaction, and 2 μl from each reaction was analyzed by gel electrophoresis. To the remainder of each sample, 2 μl of 2.5 μg/μl oyster glycogen, 5 μl of 0.5 EDTA and 200 μl of 10 mM Tris-HCl pH 7.4, 1 mM EDTA were added. The samples were phenol-chloroform extracted and isopropanol precipitated. After centrifugation the pellets were washed with 100 μl of 80% ethanol and air dried. The yield of double stranded cDNA in each of the reactions was approximately 2.5 μg.

Mung bean nuclease treatment was used to clip the single-stranded DNA of the hair-pin. Each cDNA pellet was resuspended in 15 μl of water and 2.5 μl of 10X mung bean buffer (0.3 M NaAc pH 4.6, 3 M NaCl, and 10 mM $ZnSO_4$), 2.5 μl of 10 mM DTT, 2.5 μl of 50% glycerol, and 2.5 μl of 10 U/μl mung bean nuclease (New England Biolabs, Beverly, MA) were added to each tube. The reactions were incubated at 30°C for 30 minutes and 75 μl of 10 mM Tris-HCl pH 7.4 and 1 mM EDTA was added to each tube. Two-microliter aliquots were analyzed by alkaline agarose

gel analysis. One hundred microliters of 1 M Tris-HCl pH 7.4 was added to each tube and the samples were phenol-chloroform extracted twice. The DNA was isopropanol precipitated and pelleted by centrifugation. After centrifugation, the DNA pellet was washed with 80% ethanol and air dried. The yield was approximately 2 μg of DNA from each of the two reactions.

The cDNA ends were blunted by treatment with T4 DNA polymerase. DNA from the two samples were combined after resuspension in a total volume of 24 μl of water. Four microliters of 10X T4 buffer (330 mM Tris-acetate pH 7.9, 670 mM KAc, 100 mM MgAc, and 1 mg/ml gelatin), 4 μl of 1 mM dNTP, 4 μl 50 mM DTT, and 4 μl of 1 U/μl T4 DNA-polymerase (Boehringer-Mannheim) were added to the DNA. The samples were incubated at 15°C for 1 hour. After incubation, 160 μl of 10 mM Tris-HCl pH 7.4, 1 mM EDTA was added, and the sample was phenol-chloroform extracted. The DNA was isopropanol precipitated and pelleted by centrifugation. After centrifugation the DNA was washed with 80% ethanol and air dried.

After resuspension of the DNA in 6.5 μl water, Eco RI adapters were added to the blunted DNA. One microliter of 1 μg/μl Eco RI adapter (Invitrogen, San Diego, CA Cat. # N409-20), 1 μl of 10X ligase buffer (0.5 M Tris pH 7.8 and 50 mM $MgCl_2$), 0.5 μl of 10 mM ATP, 0.5 μl of 100 mM DTT, and 1 μl of 1 U/μl T4 DNA ligase (Boehringer-Mannheim) were added to the DNA. After the sample was incubated overnight at room temperature, the ligase was heat denatured at 65°C for 15 minutes.

The Sst I cloning site encoded by the first strand primer was exposed by digestion with Sst I endonuclease. Thirty-three microliters of water, 5 μl of 10X Sst I buffer (0.5 M Tris pH 8.0, 0.1 M $MgCl_2$, and 0.5 M NaCl), and 2 μl of 5 U/μl Sst I were added to the DNA, and the samples were incubated at 37°C for 2 hours. One hundred and fifty microliters of 10 mM Tris-HCl pH 7.4, 1 mM EDTA was added, the sample was phenol-chloroform extracted, and the DNA was isopropanol precipitated.

The cDNA was chromatographed on a Sepharose CL 2B (Pharmacia LKB Biotechnology) column to size-select the cDNA and to remove free adapters. A 1.1 ml column of Sepharose CL 2B was poured into a 1 ml plastic disposable pipet and the column was washed with 50 column volumes of buffer (10 mM Tris pH 7.4 and 1 mM EDTA). The sample was applied, one-drop fractions were collected, and the DNA in the void volume was pooled. The fractionated DNA was isopropanol precipitated. After centrifugation the DNA was washed with 80% ethanol and air dried.

A Fusarium oxysporum cDNA library was established by ligating the cDNA to the vector pYcDE8' (cf. WO 90/10698) which had been digested with Eco RI and Sst I. Three hundred and ninety nanograms of vector was ligated to 400 ng of cDNA in a 80 µl ligation reaction containing 8 µl of 10 X ligase buffer, 4 µl of 10 mM ATP, 4 µl 200 mM DTT, and 1 unit of T4 DNA ligase (Boehringer-Mannheim. After overnight incubation at room temperature, 5 µg of oyster glycogen and 120 µl of 10 mM Tris-HCl and 1 mM EDTA were added and the sample was phenol-chloroform extracted. The DNA was ethanol precipitated, centrifuged, and the DNA pellet washed with 80% ethanol. After air drying, the DNA was resuspended in 3 µl of water. Thirty seven microliters of electroporation competent DH10B cells (Gibco-BRL) was added to the DNA, and electroporation was completed with a Bio-Rad Gene Pulser (Model #1652076) and Bio-Rad Pulse Controller (Model #1652098) electroporation unit (Bio-Rad Laboratories, Richmond, CA). Four milliliters of SOC (Hanahan, J. Mol. Biol. 166 (1983), 557-580) was added to the electroporated cells, and 400 µl of the cell suspension was spread on each of ten 150 mm LB amipicillin plates. After an overnight incubation, 10 ml of LB amp media was added to each plate, and the cells were scraped into the media. Clycerol stocks and plasmid preparations were made from each plate. The library background (vector without insert) was established at aproximately 1% by ligating the vector without insert and titering the number of clones after electroporation.

Screening the cDNA library

Full length cellulase cDNA clones were isolated from the Fusarium oxysporum cDNA library by hybridization to PCR generated genomic oligonucleotide probes.

The PCR-generated oligonucleotides: ZC3309, a 40-mer coding for part of the C family cellobiohydrolase, ATT ACC AAC ACC AGC GTT GAC ATC ACT GTC AGA GGG CTT C; ZC3310, a 28-mer coding for the C family endoglucanase, AAC TCC GTT GAT GAA AGG AGT GAC GTA G; and ZC3311, a 40-mer coding for the F family cellulase, CGG AGA GCA GCA GGA ACA CCA GAG GCA GGG TTC CAG CCA C, were end labeled with T_4 polynucleotide kinase and $^{32-P}$ gamma ATP. For the kinase reaction 17 picomoles of each oligonucleotide were brought up to 12.5 µl volume with deionized water. To these were added 2 µl 10 X kinase buffer (1 X: 10 mM magnesium chloride, 0.1 mM EDTA, 50 mM Tris pH 7.8), 0.5 µl 200 mM dithiothreitol, 1 µl ^{32P} gamma ATP 150 mCi/ml, Amersham), 2 µl T_4 polynucleotide kinase (10 U/µl BRL). The samples were then mixed and incubated at 37°C for 30 minutes. Oligonucleotides were separated from unincorporated nucleotides by precipitation with 180 µl TE (10 mM tris pH 8.0, 1 mM EDTA), 100 µl 7.5 M ammonium acetate, 2 µl mussel glycogen (20 mg/ml, Gibco-BRL) and 750 µl 100% ethanol. Pellets were dissolved in 200 µl distilled water. To determine the amount of radioactivity incorporated in the oligonucleotides, 10 µl of 1:1000 dilutions of oligonucleotides were read without scintillation fluid in a Beckman LS 1800 Liquid Scintillation System. Activities were: 115 million cpm for ZC3309, 86 million cpm for ZC3310, and 79 million cpm for ZC3311.

Initially, a library of 20,000 cDNA clones was probed with a mixture of each of the three oligonucleotides corresponding to the C family cellobiohydrolase, C family endoglucanase and F family cellulase clones. The cDNA library was plated out from titered glycerol stocks stored at -70°C. Four thousand clones were plated out on each of five 150 mm LB ampicillin (1000

μg/ml) plates. Lifts were taken in duplicate following standard methodology Sambrook et al., Molecular Cloning, 1989) using Biotrans 0.2 μm 137 mm filters. The filters were baked at 80°C in vacuum for 2 hours, then swirled overnight in a crystallizing dish (Pharmacia LKB Biotechnology, Alameda, CA) at 37°C in 80 ml prehybridization solution (5 X Denhardt's (1X: 0.02% Ficoll, 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumen Pentax Fraction 5 (Sigma, St. Louis, MO)) 5 X SSC (1 X: 0.15 M sodium chloride, 0.15 M sodium citrate pH 7.3)), 100 μg/ml denatured sonicated salmon sperm DNA, 50 mM sodium phosphate pH 6.8, 1 mM sodium pyrophosphate, 100 μM ATP, 20% formamide, 1% sodium dodecyl sulfate) (Ulrich et al. EMBO J. 3 (1984), 361-364).

Prehybridized filters were probed by adding them one at a time into a crystallizing dish with 80 ml prehybridization solution with 80 million cpm ZC3309, 86 million cmp ZC3310 and 79 million cpm ZC3311 and then swirled overnight at 37°C. Filters were then washed to high stringency. The probed filters were washed with three 400 ml volumes of low stringency wash solution (2 X SSC, 0.1% SDS) at room temperature in the crystallizing dish, then with four 1-liter volumes in a plastic box. A further wash for 20 minutes at 68°C with tetramethylammonium chloride wash solution (TMACL: 3 M tetramethylammonium chloride, 50 mM Tris-HCl pH 8.0, 2 mM EDTA, 1 g/l SDS) (Wood et al., Proc. Natl. Acad. Sci. 82 (1985)) provided a high stringency wash for the 28-mer ZC3310 independent of its base composition 1585-1588). The filters were then blotted dry, mounted on Whatman 3MM paper and covered with plastic wrap for autoradiography. They were exposed overnight at -70°C with intensifying screens and Kodak XAR-5 film.

Two putative positives appeared on duplicate filters. The corresponding areas on the plates with colonies were picked into 1 ml of 1X polymerase chain reaction (PCR) buffer (100 mM Tris HCl pH 8.3, 500 mM KCl, 15 mM MgCl, 0.1% gelatin; Perkin

Elmer Cetus) and plated out at five tenfold dilutions on 100 mm LB plates with 70 μg/ml ampicillin. These plates were grown at 37°C overnight. Two dilutions of each putative clone were chosen for rescreening as outlined above. One isolated clone, pZFH196 was found. This was grown up overnight in 10 ml 2X YT broth (per liter: 16 g bacto-tryptone, 10 g bacto-yeast extract, 10 g NaCl). Twenty three micrograms of DNA were purfified by the rapid boiling method (Holmes and Quigley, Anal. Biochem. 114 (1981), 193-197). From restriction analysis the clone was found to be approximately 2,000 base pairs in length. Sequence analysis showed it to contain a fragment homologous to the C family cellobiohydrolase fragment cloned by PCR.

In an attempt to isolate additional cellulase cDNA clones, a cDNA library of 2 million clones was plated out on 20 150 mm LB plates (100 μg/ml ampicillin) containing approximately 100,000 cDNA clones. Lifts were taken in duplicate as in the first screening attempt. This library was screened with oligonucleotides corresponding to the three cellulase species as described above except that the hybridization was carried out with formamide in the prehybridization buffer and at a temperature of 30°C. Washing with TMACL was carried out twice for 20 minutes at 67°C. Between 8 and 20 signals were found on duplicate filters of each of the 20 plates. Fifteen plugs were taken from the first plate with the large end of a pasteur pipet into 1 ml 1 X PCR buffer (Perkin-Elmer Cetus). PCR was carried out on the bacterial plugs with three separate oligonucleotide mixtures. Each mixture contained the vector specific oligonucleotide ZC2847 and additionally, a different cellulase specific oligonucleotide (ZC3309, ZC3310 or ZC3311) within each mixture. Amplitaq polymerase (Perkin-Elmer Cetus) was used with Pharmacia Ultrapure dNTP and following Perkin Elmer Cetus procedures. Sixteen picomoles of each primer were used in 40 μl reaction volumes. Twenty microliters of cells in 1 X PCR buffer were added to 20 μl mastermix which contained everything needed for PCR except for DNA. After an initial 1

minute 45 second denaturation at 94°C 28 cycles of: 45 seconds at 94°C, 1 minute at 45°C and 2 minutes at 72°C with a final extension of 10 minutes at 72°C were employed in a Perkin Elmer thermocycler. Ten of the 15 plugs yielded a band when primed with the C family specific oligonucleotide ZC3309 and ZC2847. The other mixtures gave no specific products. Five plugs which produced the largest bands by PCR, therefore possibly being full length C family cellobiohydrolases, along with the 5 plugs which did not produce PCR bands, were plated out at five 10 fold dilutions onto 100 mm LB plates with 70 μg/ml ampicillin and grown overnight. Duplicate lifts were taken of two ten fold dilutions each. Prehydridization and hybridization were carried out as described above with a mixture of the 3 oligonucleotides. Isolated clones were found on all 10 of the platings. These were picked from the dilution plates with a toothpick for single colony isolation on 100 mm LB plates with 70 μg/ml ampicillin. PCR was carried out on isolated bacterial colonies with 2 oligonucleotides specific for the C family cellobiohydrolase (ZC3409 (CCG TTC TGG ACG TAC AGA) and ZC3411 (TGA TGT CAA GTT CAT CAA)). Conditions were identical to those described above except for using 10 picomoles of each primer in 25 μl reaction volumes. Colonies were added by toothpick into PCR tubes with 25 μl mastermix before cycling. Five of the 10 gave strong bands of the size expected for a C family cellobiohydrolase. Isolated colonies were then grown up in 20 ml of Terrific Broth (Sambrook et al., op. cit., A2) and DNA was isolated by the rapid boiling method. The clones were partially sequenced by Sanger dideoxy sequencing. From sequence analysis the 5 clones which did not give bands specific for a C family cellobiohydrolase by PCR were shown to be F family cellulase clones.

In order to clone the C family endoglucanase, the cDNA library of 2 million clones was rescreened with only ZC3310. Conditions of prehydridization and hybridization were like those used above. Filters were hybridized for 10 hours at 30°C with one million CPM endlabeled ZC3310 per ml prehybridization solution

without formamide. Washing with TMACL was carried out 2 times for 20 minutes at 60°C. Seven weak signals were found on duplicate filters. Plugs were picked with the large end of a pipet into 1 ml LB broth. These were each plated out in 5 10 fold dilutions on 100 mm LB plates with 70 μg/ml ampicillin. Duplicate lifts were taken of 2 dilutions each and were processed as described above. Prehybridization, hybridization, and washing were carried out as for the first level of screening. Three isolated clones were identified and streaked out for single colony hybridization. Isolates were grown overnight in 50 ml of Terrific Broth (per liter: 12 g tryptone, 24 g yeast extract, 4 ml glycerol, autoclaved, and 100 ml of 0.17 M KH_2PO_4, 0.72 M K_2HPO_4 (Sambrook et al., op. cit., A2) and DNA was isolated by alkaline lysis and PEG precipitation by standard methods (Maniatis 1989, 1.38-1.41). From restriction analysis, one clone (pZFH223) was longer than the others and was chosen for complete sequencing. Sequence analysis showed it to contain the PCR fragment cloned initially.

DNA sequence analysis

The cDNAs were sequenced in the yeast expression vector pYCDE8'. The dideoxy chain termination method (F. Sanger et al., Proc. Natl. Acad. Sci. USA 74, 1977, pp. 5463-5467) using @35-S dATP from New England Nuclear (cf. M.D. Biggin et al., Proc. Natl. Acad. Sci. USA 80, 1983, pp. 3963-3965) was used for all sequencing reactions. The reactions were catalysed by modified t7 DNA polymerase from Pharmacia (cf. S. Tabor and C.C. Richardson, Proc. Natl. Acad. Sci. USA 84, 1987, pp. 4767-4771) and were primed with an oligonucleotide complementary to the ADH1 promoter (ZC996: ATT GTT CTC GTT CCC TTT CTT), complementary to the CYC1 terminator (ZC3635: TGT ACG CAT GTA ACA TTA) or with oligonucleotides complementary to the DNA of interest. Double stranded templates were denatured with NaOH (E.Y. Chen and P.H. Seeburg, DNA 4, 1985, pp. 165-170) prior to hybridizing with a sequencing oligonucleotide. Oligonucleotides were synthesized on an Applied Biosystems Model 380A DNA

synthesizer. The oligonucleotides used for the sequencing reactions are listed in the sequencing oligonucleotide table below:

C-family cellobiohydrolase sequencing primers

ZC3411	TGA TGT CAA GTT CAT CAA
ZC3408	TCT GTA CGT CCA GAA CGG
ZC3407	ATG ACT TCT CTA AGA AGG
ZC3406	TCC AAC ATC AAG TTC GGT
ZC3410	AGG CCA ACT CCA TCT GAA
ZC3309	ATT ACC AAC ACC AGC GTT GAC ATC ACT GTC AGA GGG CTC C
ZC3409	CCG TTC TGG ACG TAC AGA

F-family cellulase specific sequencing primers

ZC3413	CCA TCG ACG GTA TTG GAT
ZC3311	CGG AGA GCA GCA GGA ACA CCA GAG GCA GGG TTC CAG CCA C
ZC3412	GAG GGT AGA GCG ATC GTT

C-family endoglucanase specific sequencing primers

ZC3739	TGA TCT CAT CGA GCT GCA CC
ZC3684	GTG ATG CTC AGT GCT ACG TC
ZC3310	AAC TCC GTT GAT GAA AGG AGT GAC GTA G
ZC3750	TCC AAT AGC TTC CCA GCA AG
ZC3683	TGT CCC TTG ATG TTG CCA AC

The DNA sequences of the full-length cDNA clones, as well as the derived amino acid sequences, are shown in the appended Figs. 11 (C-family cellobiohydrolase), 12 (F-family cellulase) and 13 (C-family endoglucanase).

Example 4

Isolation of endoglucanase EGI gene from H. insolens

The cDNA library described in example 1 was also screened with a 35 bp oligonucleotide probe in the antisense configuration with the sequence:

NOR-770: 5' GCTTCGCCCATGCCTTGGGTGGCGCCGAGTTCCAT 3'

The sequence was derived from the amino acid sequence of an alcalase fragment of EGI purified from H. insolens, using our knowledge of codon bias in this organism. Complete clones of 1.6 kb contained the entire coding sequence of 1.3 kb as shown in Fig. 14A-E. The probe sequence NOR-770 is located at Met_{344}-Ala_{355}.

Construction of expression plasmids of EGI (full length) and EGI' (truncated)

The EGI gene still containing the poly-A tail was inserted into an A. oryzae expression plasmid as outlined in Fig 2. The coding region of EGI was cut out from the NcoI-site in the initiating Met-codon to the Bam H1-site downstream of the poly-A region as a 1450 bp fragment from pHW480. This was ligated to a 3.6 kb NcoI-NarI fragment from pSX224 (Fig. 1) containing the TAKA promoter and most of pUC19, and to a 960 bp NarI-BamHI fragment containing the remaining part with the AMG-terminator. The 960 bp fragment was taken from p960 which is equivalent to p777 (described in EP 238,023) except for the inserted gene. The resulting expression plasmid is termed pHW485.

The expression plasmid pHW704 with the full length EGI gene without poly A tail is shown in Fig. 3. From the BstEII site 1300 bp downstream of the NcoI-site was inserted a 102 bp BstEII-BamHI linker (2645/2646) ligated to BglII-site in the vector. The linker contains the coding region downstream of BstEII-site with 2 stop codons at the end and a PvuI-site near the C-terminal to be used for addition of CBD and B-regions.

Expression plasmid pHW697 with the truncated EGI' gene was constructed similarly using a BstEII-BamHI linker (2492/2493) of 69 bp. In this linker we introduced a PstI-site altering Val_{421} to Leu_{421} and the last 13 amino acids of the coding region: $K_{423}PKPKPGHGPRSD_{435}$ were eliminated. The short tail with the rather unusual sequence was cut off to give EGI' a C-terminal corresponding to the one found in T. reesei EGI just upstream of the A and B-region.

Construction of an expression plasmid of EGI' with CBD and B region from a ~ 43 kD endoglucanase added C-terminally

The ~ 43 kD endoglucanase of H. isolens described in DK patent application No. 736/91 has shown good washing performance. Besides the catalytic domain, 43 kD cellulase has a C-terminal CBD and B region which has been transferred to EGI' which does not have any CBD or B region itself. The construction was done in 2 steps, as outlined in Fig. 4. The PstI-HincII linker (028/030 M) intended to connect the C-terminal of EGI' to the B-region of 43 kD cellulase, was subcloned in pUC19 PstI-EcoRI with C-terminal Hinc2-EcoRI 100 bp fragment from 43 kD cellulase gene in pSX320 (Fig 5; as described in DK 736/91). From the subclone pHW767 the CBD and B-region was cut out as a 250 bp PstI-BglII fragment and ligated to pHW485 (Fig. 2) BstEII-BglII fragment of 5.7 kb and to the remaining BstEII-PstI fragment of 55 bp from pHW697 (Fig. 3). The resulting expression plasmid pHW768 has the ~ 43 kD endoglucanase CBD and B region added to Gln_{422} of EGI'.

Construction of an expression plasmid of EGI with the CBD and B region from ~ 43 kD endoglucanase added C-terminally

This plasmid was constructed in a similar way as pHW768 except that, in this case, the C-terminal linker yielded the complete sequence of EGI. Fig. 6 shows the procedure in 3 steps. The

PvuI-HincII linker (040 M/041 M) was subcloned in pUC18 to give pHW775, into which a HincII-EcoRI 1000 bp fragment from pSX 320 (Fig. 5) was inserted to give pHW776. From this the CBD and B region was cut out as a 250 bp PvuI-BglII fragment and ligated to 5.7 kb BstEII-BglII fragment from pHW485 (Fig. 2) and 90 bp BstEII-Pvu1 fragment from pHW704 (Fig. 3). The resulting expression plasmid pHW777 contains the ~ 43 kD endoglucanase CBD and B region added to Asp_{435} in the complete EGI sequence.

Expression in A. oryzae of EGI and EGI' with and without the CBD and B region from ~ 43 kD endoglucanase

The expression plasmids pHW485, pHW704, pHW697, pHW768 and pHW777 were transformed into A. oryzae IFO 4177 as described in example 2. Supernatants from transformants grown in YPD medium as described were analyzed by SDS-PAGE, where the native EGI has an apparent Mw of 53 kD. EGI' looks slightly smaller as expected, and the species with the added CBD and B region are increased in molecular weight corresponding to the size of the CBD and B region with some carbohydrate added. A polyclonal antibody AS169 raised against the ~ 43 kD endoglucanase recognizes EGI and EGI' only when the ~ 43 kD CBD and B region are added, while all 4 species are recognized by a polyclonal antibody AS78 raised against a cellulase preparation from H. insolens. All 4 species have endoglucanase activity as measured on soluble cellulose in the form of carboxy methyl cellulose.

Linkers

2492/2493: BstE2-Pst1-BamH1

```
5'      GTCACCTACACCAACCTCCGCTGGGGCGAG
3'           GATGTGGTTGGAGGCGACCCCGCTC

        ATCGGCTCGACCTACCAGGAGCTGCAGTAGTAA
        TAGCCGAGCTGGATGGTCCTCGACGTCATCATT

        TGATAG          3'          69 bp
        ACTATCCTAG      5'          68 bp
```

2645/2646: BstE2-Xma1-PvuI-BamH1

```
5'      GTCACCTACACCAACCTCCGCTGGGGCGAGATCGGC
3'           GATGTGGTTGGAGGCGACCCCGCTCTAGCCG

        TCGACCTACCAGGAGGTTCAGAAGCCTAAGCCCAAG
        AGCTGGATGGTCCTCCAAGTCTTCGGATTCGGGTTC

        CCCGGGCACGGCCCCCGATCGGACTAATAG        3'   102 bp
        GGGCCCGTGCCGGGGGCTAGCCTGATTATCCTAG    5'   101 bp
```

028 M/030 M: PstI-HincII

```
5'          GTCCAGCAGCACCAGCTCTCCGGTC         3'    25 bp
3'     ACGTCAGGTCGTCGTGGTCGAGAGGCCAG          5'    29 bp
```

040 M/041 M: PvuI-HincII

```
5'        CGTCCAGCAGCACCAGCTCTCCGGTC         3'    26 bp
3'     TAGCAGGTCGTCGTGGTCGAGAGGCCAG          5'    28 bp
```

Example 5

- 43 kD endoglucanase with different CBDs and B-regions:

In order to test the influence on the - 43 kD endoglucanase of the different CBDs and B regions from the A region clones we have substituted the original CBD and B region from - 43 kD with the other C-terminal CBDs and B regions, i.e. A-1, A-8, A-9, A-11, and A-19 (cf. Example 1). In order to test the concept we have also made a construction where the 43 kD B region has been deleted.

Fragments:

All fragments were made by PCR amplification using a Perkin-Elmer/Cetus DNA Amplification System following the manufacturers instructions.

1) A PCR fragment was made which covers the DNA from 56 bp upstream of the Bam HI site on pSX 320 (Fig. 5) to 717 bp within the coding region of the ~43 kD endoglucanase gene and at the same time introduces a Kpn I site at pos. 708 and a Sma I site at pos. 702 in the coding region which is at the very beginning of the B region. This PCR fragment was made with the primers NOR 1542 and NOR 3010 (see list of oligonucleotides below).

2) A PCR fragment was made which includes the CBD and B region of A-1 introducing a Kpn I site at the very beginning of the B region in frame with the Kpn I site introduced in 1) and introducing a Xho I site downstream of the coding region of the gene. Primers used: NOR 3012 upstream and NOR 3011 downstream.

3) As 2) except that the fragment covered the CBD and B region of A-8 and the Xho I site in the expression vector downstream of gene. Primers: NOR 3017 and NOR 2516.

4) As 2) but with primers NOR 3016 and NOR 3015 covering the CBD and B region from A-9.

5) As 3) but with primers NOR 3021 and NOR 2516 covering the CBD and B region from A-11.

6) As 2) but with primers NOR 3032 and NOR 3022 covering the CBD and B region from A-19.

7) A PCR fragment which includes the CBD from ~ 43 kD endoglucanase and the Xho I site downstream from the gene on pSX 320 introducing a Pvu II site at the very end of the B region.
Primers: NOR3023 and NOR2516.

Combinations:

1) + 2) inserted as Bam HI - Kpn I and Kpn I - Xho I into pToC 68 (described in DK736/91) Bam HI - Xho I, thus coding for the 43 kD core enzyme with the CBD and B region from A-1.

1) + 3): Like above giving a 43 kD enzyme with the A-8 CBD/B region.

1) + 4): As above, but with the A-9 CBD and B region.

1) + 5): As above, but with the A-11 CBD and B region.

1) + 6): As above, but with the A-19 CBD and B region.

1) + 7) inserted as Bam HI - Sma I and Pvu II - Xho I into pToC 68 Bam HI - Xho I, thus coding for the 43 kD enzyme without the B region.

Oligonucleotides:

NOR 1542: 5' - CGACAACATCACATCAAGCTCTCC - 3'

NOR 2516: 5' - CCATCCTTTAACTATAGCGA - 3'

NOR 3010: 5' - GCTGGTGCTGGTACCCGGGATCTGGACGGCAGGG - 3'
Kpn Sma

NOR 3011: 5' - GCATCGGTACCGGCGGCGGCTCCACTGGCG - 3'
Kpn

NOR 3012: 5' - CTCACTCCATCTCGAGTCTTTCAATTTACA - 3'
Xho

NOR 3015: 5' - CTTTTCTCGAGTCCCTTAGTTCAAGCACTGC - 3'
Xho

NOR 3016: 5' - TGACCGGTACCGGCGGCGGCAACACCAACC - 3'
Kpn

NOR 3017: 5' - TCACCGGTACCGGCGGTGGAAGCAACAATG - 3'
Kpn

NOR 3021: 5' - TCTTCGGTACCAGCGGCAACAGCGGCGGCG - 3'
Kpn

NOR 3022: 5' - CGCTGGGTACCAACAACAATCCTCAGCAGG -3'
Kpn

NOR 3023: 5' - CTCCCAGCAGCTGCACTGCTGAGAGGTGGG - 3'
Pvu II

NOR 3032: 5' - CGGCCTCGAGACCTTACAGGCACTGCGAGT - 3'
Xho

Example 6

Fusion of a bacterial catalytic domain to a fungal CBD

The endoglucanase Endo 1 produced by Bacillus lautus NCIMB 40250 (described in PCT/DK91/00013) consists of a catalytic domain (core) (Ala(32) - Val(555)) and a C terminal cellulose binding domain (CBD) (Gln556 - Pro700) homologous to the CBD of a B. subtilis endoglucanase (R.M. MacKay et al. 1986. Nucleic Acids Res. 14, 9159-70). The CBD is proteolytically cleaved off when the enzyme is expressed in B. subtilis or E. coli generating a CMC degrading core enzyme. In this example this core protein was fused with the B region and CBD of the ~ 43 kD endoglucanase from Humicola insolens (described in DK 736/91).

Construction of the fusion.

The plasmid pCaHj 170 containing the cDNA gene encoding the ~ 43 kD endoglucanase was constructed as shown in Fig. 7. pCaHj 170 was digested with Xho II and Sal I. The 223 bp Xho II - Sal I fragment was isolated and ligated into pUC 19 (Yanisch-Perron et al. 1985. Gene 33, 103-119) digested with BamH I and Sal I. The BamH I site was regenerated by this Xho II-BamH I ligation. The resulting plasmid, IM 2, was digested with Eco R1 and BamH I and ligated with the linker NOR 3045 - NOR 3046:

NOR 3045 5' AATTCCGCGGAACGATATCTCCGA 3'
NOR 3046 3' GGCGCCTTGCTATAGAGGCTCTAG 5'
EcoR I EcoR V Mbo I
Sac II

The resulting plasmid, IM 3, was digested with EcoR V and SacII and ligated to the 445 bp Hinc II - Sac II pPL 517 fragment. pPL 517 contains the entire Bacillus Endo 1 gene (PCT/DK91/00013). The product of this ligation was termed IM 4. In order to replace the Bacillus signal peptide of Endo 1 with the fungal signal peptide from the 43 kdal endoglucanase four PCR primers were designed for "Splicing by Overlap Extension" (SOE) fusion (R M Horton et al.(1989):Gene, 77, 61-68). The 43 kD signal sequence was amplified from the plasmid pCaHj 109 (DK 736/91) introducing a Bcl I site in the 5'end and a 21 bp homology to the Bacillus endo 1 gene in the 3' end using the 5' primer NOR 3270 and the 3' primer NOR 3275. The part of the Endo I gene 5' to the unique Sac II site was amplified using the 5' primer NOR 3276 introducing a 21 bp homology to the 43 kdal gene and the 3' primer NOR 3271 covering the Sac II site. The two PCR framents were mixed, melted, annealed and filled up with the taq polymerase (Fig. 9). The resulting hybrid was amplified using the primers NOR 3270 and NOR 3271. The hybrid fragment was digested with Bcl 1 and SacII and ligated to the 676 bp Sac II - Sal I fragment from IM 4 and the Aspergillus expression vector pToc 68 (DK 736/91) digested with BamH I. The product of this ligation, pCaHj 180 (Fig. 10), contained an open reading frame encoding the 43 kD signal peptide and the first four N terminal aminoacids of the mature ~ 43 kD endoglucanase (Met(1)-Arg(25) fused to the core of Endo 1 (Ser(34)-Val(549)) followed by the peptide Ile-Ser-Glu (encoded by the linker) fused to the 43 kD B region and CBD (Ile(233)-Leu(285). pCaHj 180 was used to transform Aspergillus oryzae IFO 4177 using selection on acetamide by cotransformation with pToC 90 (cf. DK 736/91) as described in published EP patent application No. 238 023.

NOR 3270 5' TTGAATTCTGATCAAGATGCGTTCCTCCC 3'
NOR 3275 5' AATGGTGAAAGTGACATCACTCCTGCCATCAGCGGCAAGGGC 3'
NOR 3276 5' GCCCTTGCCGCTGATGGCAGGAGTGATGTCACTTTCACCATT 3'
NOR 3271 5' AGCGCGTCCGCGGTAGCTATG 3'

The sequence of the Endo 1 core and the ~ 43 kD CBD and B region is shown in the appended Fig. 15A-D.

CLAIMS

1. A cellulose- or hemicellulose-degrading enzyme which is derivable from a fungus other than Trichoderma or Phanerochaete, and which comprises a carbohydrate binding domain homologous to a terminal A region of Trichoderma reesei cellulases, which carbohydrate binding domain comprises the following amino acid sequence

1 10
Xaa Xaa Gln Cys Gly Gly Xaa Xaa Xaa Xaa Gly Xaa Xaa Xaa Cys Xaa

20 30
Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Asn Xaa Xaa Tyr Xaa Gln Cys Xaa
-
Xaa
-

or a subsequence thereof capable of effecting binding of the enzyme to an insoluble cellulosic or hemicellulosic substrate.

2. An enzyme according to claim 1, which is derivable from a strain of Humicola, Fusarium or Myceliopthora.

3. An enzyme according to claim 1, wherein the variations in the amino acid sequence shown in claim 1 are selected as follows
in position 1, the amino acid is Trp or Tyr;
in position 2, the amino acid is Gly or Ala;
in position 7, the amino acid is Gln, Ile or Asn;
in position 8, the amino acid is Gly or Asn;
in position 9, the amino acid is Trp, Phe or Tyr;
in position 10, the amino acid is Ser, Asn, Thr or Gln;
in position 12, the amino acid is Pro, Ala or Cys;
in position 13, the amino acid is Thr, Arg or Lys;
in position 14, the amino acid is Thr, Cys or Asn;
in position 18, the amino acid is Gly or Pro;
in position 19, the amino acid (if present) is Ser, Thr, Phe, Leu or Ala;
in position 20, the amino acid is Thr or Lys:

in position 24, the amino acid is Gln or Ile;
in position 26, the amino acid is Gln, Asp or Ala;
in position 27, the amino acid is Trp, Phe or Tyr;
in position 29, the amino acid is Ser, His or Ala; and/or
in position 32, the amino acid is Leu, Ile, Gln, Val or Thr.

4. An enzyme according to claim 3, wherein the carbohydrate binding domain comprises the following amino acid sequence

Trp Gly Gln Cys Gly Gly Gln Gly Trp Asn Gly Pro Thr Cys Cys Glu Ala Gly Thr Thr Cys Arg Gln Gln Asn Gln Trp Tyr Ser Gln Cys Leu;

Trp Gly Gln Cys Gly Gly Ile Gly Trp Asn Gly Pro Thr Thr Cys Val Ser Gly Ala Thr Cys Thr Lys Ile Asn Asp Trp Tyr His Gln Cys Leu;

Trp Gly Gln Cys Gly Gly Ile Gly Phe Asn Gly Pro Thr Cys Cys Gln Ser Gly Ser Thr Cys Val Lys Gln Asn Asp Trp Tyr Ser Gln Cys Leu;

Trp Gly Gln Cys Gly Gly Asn Gly Tyr Ser Gly Pro Thr Thr Cys Ala Glu Gly - Thr Cys Lys Lys Gln Asn Asp Trp Tyr Ser Gln Cys Thr Pro;

Trp Gly Gln Cys Gly Gly Gln Gly Trp Gln Gly Pro Thr Cys Cys Ser Gln Gly - Thr Cys Arg Ala Gln Asn Gln Trp Tyr Ser Gln Cys Leu Asn;

Trp Gly Gln Cys Gly Gly Gln Gly Tyr Ser Gly Cys Thr Asn Cys Glu Ala Gly Ser Thr Cys Arg Gln Gln Asn Ala Tyr Tyr Ser Gln Cys Ile;

Trp Gly Gln Cys Gly Gly Gln Gly Tyr Ser Gly Cys Arg Asn Cys Glu Ser Gly Ser Thr Cys Arg Ala Gln Asn Asp Trp Tyr Ser Gln Cys Leu;

Trp Ala Gln Cys Gly Gly Asn Gly Trp Ser Gly Cys Thr Thr Cys Val Ala Gly Ser Thr Cys Thr Lys Ile Asn Asp Trp Tyr His Gln Cys Leu;

Trp Gly Gln Cys Gly Gly Gln Asn Tyr Ser Gly Pro Thr Thr Cys Lys Ser Pro Phe Thr Cys Lys Lys Ile Asn Asp Phe Tyr Ser Gln Cys Gln; or

Trp Gly Gln Cys Gly Gly Asn Gly Trp Thr Gly Ala Thr Thr Cys Ala Ser Gly Leu Lys Cys Glu Lys Ile Asn Asp Trp Tyr Tyr Gln Cys Val

5. An enzyme according to any of claims 1-4, which further comprises an amino acid sequence which defines a linking B region connecting the carbohydrate binding domain to the catalytically active domain of the enzyme.

6. An enzyme according to claim 5, wherein the linking B region is one which is enriched in the amino acids glycine and/or asparagine and/or proline and/or serine and/or threonine and/or glutamine.

7. An enzyme according to claim 6, wherein one or more of said amino acids appear in short, repetitive units.

8. An enzyme according to any of claims 1-7, which comprises a carbohydrate binding domain derived from one naturally occurring cellulose- or hemicellulose-degrading enzyme, an amino acid sequence defining a linking B region, which amino acid sequence is derived from another naturally occurring cellulose- or hemicellulose-degrading enzyme, as well as a catalytically active domain derived from the enzyme supplying either the carbohydrate binding domain or B region or from a third enzyme.

9. An enzyme according to claim 8, wherein the catalytically active domain is derived from an enzyme which does not, in nature, comprise a carbohydrate binding domain or B region.

10. An enzyme according to any of claims 1-9 which is a cellulase, e.g. an endoglucanase, cellobiohydrolase or β-glucosidase.

11. A DNA construct which comprises a DNA sequence encoding an enzyme according to any of claims 1-10.

12. An expression vector which carries an inserted DNA construct according to claim 11.

13. A cell which is transformed with a DNA construct according to claim 11 or with an expression vector according to claim 12.

14. A cell according to claim 13 which is a fungal cell, e.g. belonging to a strain of Aspergillus, e.g. Aspergillus niger or Aspergillus oryzae, or a yeast cell, e.g. belonging to a strain of Saccharomyces, such as Saccharomyces cerevisiae.

15. A method of producing an enzyme according to any of claims 1-10, wherein a cell according to claim 13 or 14 is cultured under conditions conducive to the production of the enzyme, and the enzyme is subsequently recovered from the culture.

16. An agent for degrading cellulose or hemicellulose, the agent comprising an enzyme according to any of claims 1-10.

17. An agent according to claim 16 comprising a combination of two or more enzymes according to any of claims 1-10, or a combination of one or more enzymes according to any of claims 1-10 with one or more other enzymes with cellulose- or hemicellulose-degrading activity.

18. A carbohydrate binding domain homologous to a terminal A region of Trichoderma reesei cellulases, which carbohydrate binding domain comprises the following amino acid sequence

1 10
Xaa Xaa Gln Cys Gly Gly Xaa Xaa Xaa Xaa Gly Xaa Xaa Xaa Cys Xaa

20 30
Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Asn Xaa Xaa Tyr Xaa Gln Cys Xaa

Xaa

or a subsequence thereof capable of effecting binding of a protein to an insoluble cellulosic or hemicellulosic substrate.

19. A carbohydrate binding domain according to claim 18, wherein the variations in the amino acid sequence shown in claim 18 are selected as follows

in position 1, the amino acid is Trp or Tyr;
in position 2, the amino acid is Gly or Ala;
in position 7, the amino acid is Gln, Ile or Asn;
in position 8, the amino acid is Gly or Asn;
in position 9, the amino acid is Trp, Phe or Tyr;
in position 10, the amino acid is Ser, Asn, Thr or Gln;
in position 12, the amino acid is Pro, Ala or Cys;
in position 13, the amino acid is Thr, Arg or Lys;
in position 14, the amino acid is Thr, Cys or Asn;
in position 18, the amino acid is Gly or Pro;
in position 19, the amino acid (if present) is Ser, Thr, Phe, Leu or Ala;
in position 20, the amino acid is Thr or Lys:
in position 24, the amino acid is Gln or Ile;
in position 26, the amino acid is Gln, Asp or Ala;
in position 27, the amino acid is Trp, Phe or Tyr;
in position 29, the amino acid is Ser, His or Tyr; and/or
in position 32, the amino acid is Leu, Ile, Gln, Val or Thr.

20. A carbohydrate binding domain according to claim 19, wherein the carbohydrate binding domain comprises the following amino acid sequence

Trp Gly Gln Cys Gly Gly Gln Gly Trp Asn Gly Pro Thr Cys Cys Glu Ala Gly Thr Thr Cys Arg Gln Gln Asn Gln Trp Tyr Ser Gln Cys Leu;

Trp Gly Gln Cys Gly Gly Ile Gly Trp Asn Gly Pro Thr Thr Cys Val Ser Gly Ala Thr Cys Thr Lys Ile Asn Asp Trp Tyr His Gln Cys Leu;

Trp Gly Gln Cys Gly Gly Ile Gly Phe Asn Gly Pro Thr Cys Cys Gln Ser Gly Ser Thr Cys Val Lys Gln Asn Asp Trp Tyr Ser Gln Cys Leu;

Trp Gly Gln Cys Gly Gly Asn Gly Tyr Ser Gly Pro Thr Thr Cys Ala Glu Gly - Thr Cys Lys Lys Gln Asn Asp Trp Tyr Ser Gln Cys Thr Pro;

Trp Gly Gln Cys Gly Gly Gln Gly Trp Gln Gly Pro Thr Cys Cys Ser Gln Gly - Thr Cys Arg Ala Gln Asn Gln Trp Tyr Ser Gln Cys Leu Asn;

Trp Gly Gln Cys Gly Gly Gln Gly Tyr Ser Gly Cys Thr Asn Cys Glu Ala Gly Ser Thr Cys Arg Gln Gln Asn Ala Tyr Tyr Ser Gln Cys Ile;

Trp Gly Gln Cys Gly Gly Gln Gly Tyr Ser Gly Cys Arg Asn Cys Glu Ser Gly Ser Thr Cys Arg Ala Gln Asn Asp Trp Tyr Ser Gln Cys Leu;

Trp Ala Gln Cys Gly Gly Asn Gly Trp Ser Gly Cys Thr Thr Cys Val Ala Gly Ser Thr Cys Thr Lys Ile Asn Asp Trp Tyr His Gln Cys Leu;

Trp Gly Gln Cys Gly Gly Gln Asn Tyr Ser Gly Pro Thr Thr Cys Lys Ser Pro Phe Thr Cys Lys Lys Ile Asn Asp Phe Tyr Ser Gln Cys Gln; or

Trp Gly Gln Cys Gly Gly Asn Gly Trp Thr Gly Ala Thr Thr Cys Ala Ser Gly Leu Lys Cys Glu Lys Ile Asn Asp Trp Tyr Tyr Gln Cys Val

21. A linking B region derived from a cellulose- or hemicellulose-degrading enzyme, said region comprising an amino acid sequence enriched in the amino acids glycine and/or asparagine and/or proline and/or serine and/or threonine and/or glutamine.

22. A B region according to claim 21, wherein one or more of said amino acids appear in short, repetitive units.

23. A B region according to claim 21 or 22, which comprises the following amino acid sequence

Ala Arg Thr Asn Val Gly Gly Gly Ser Thr Gly Gly Gly Asn Asn Gly Gly Gly Asn Asn Gly Gly Asn Pro Gly Gly Asn Pro Gly Gly Asn Pro Gly Gly Asn Pro Gly Gly Asn Pro Gly Gly Asn Pro Gly Gly Asn Cys Ser Pro Leu;

Pro Gly Gly Asn Asn Asn Asn Pro Pro Pro Ala Thr Thr Ser Gln Trp Thr Pro Pro Pro Ala Gln Thr Ser Ser Asn Pro Pro Pro Thr Gly Gly Gly Gly Gly Asn Thr Leu His Glu Lys;

Gly Gly Ser Asn Asn Gly Gly Gly Asn Asn Asn Gly Gly Gly Asn Asn Asn Gly Gly Gly Gly Asn Asn Asn Gly Gly Gly Asn Asn Asn Gly Gly Gly Asn Thr Gly Gly Gly Ser Ala Pro Leu;

Val Phe Thr Cys Ser Gly Asn Ser Gly Gly Gly Ser Asn Pro Ser Asn Pro Asn Pro Pro Thr Pro Thr Thr Phe Ile Thr Gln Val Pro Asn Pro Thr Pro Val Ser Pro Pro Thr Cys Thr Val Ala Lys;

Pro Ala Leu Trp Pro Asn Asn Asn Pro Gln Gln Gly Asn Pro Asn Gln Gly Gly Asn Asn Gly Gly Gly Asn Gln Gly Gly Gly Asn Gly Gly Cys Thr Val Pro Lys;

Pro Gly Ser Gln Val Thr Thr Ser Thr Thr Ser Ser Ser Ser Thr Thr Ser Arg Ala Thr Ser Thr Thr Ser Ala Gly Gly Val Thr Ser Ile Thr Thr Ser Pro Thr Arg Thr Val Thr Ile Pro Gly Gly Ala Ser Thr Thr Ala Ser Tyr Asn;

Glu Ser Gly Gly Gly Asn Thr Asn Pro Thr Asn Pro Thr Asn Pro Thr Asn Pro Thr Asn Pro Thr Asn Pro Trp Asn Pro Gly Asn Pro Thr Asn Pro Gly Asn Pro Gly Gly Gly Asn Gly Gly Asn Gly Gly Asn Cys Ser Pro Leu; or

Pro Ala Val Gln Ile Pro Ser Ser Ser Thr Ser Ser Pro Val Asn Gln Pro Thr Ser Thr Ser Thr Thr Ser Thr Ser Thr Thr Ser Ser Pro Pro Val Gln Pro Thr Thr Pro Ser Gly Cys Thr Ala Glu Arg

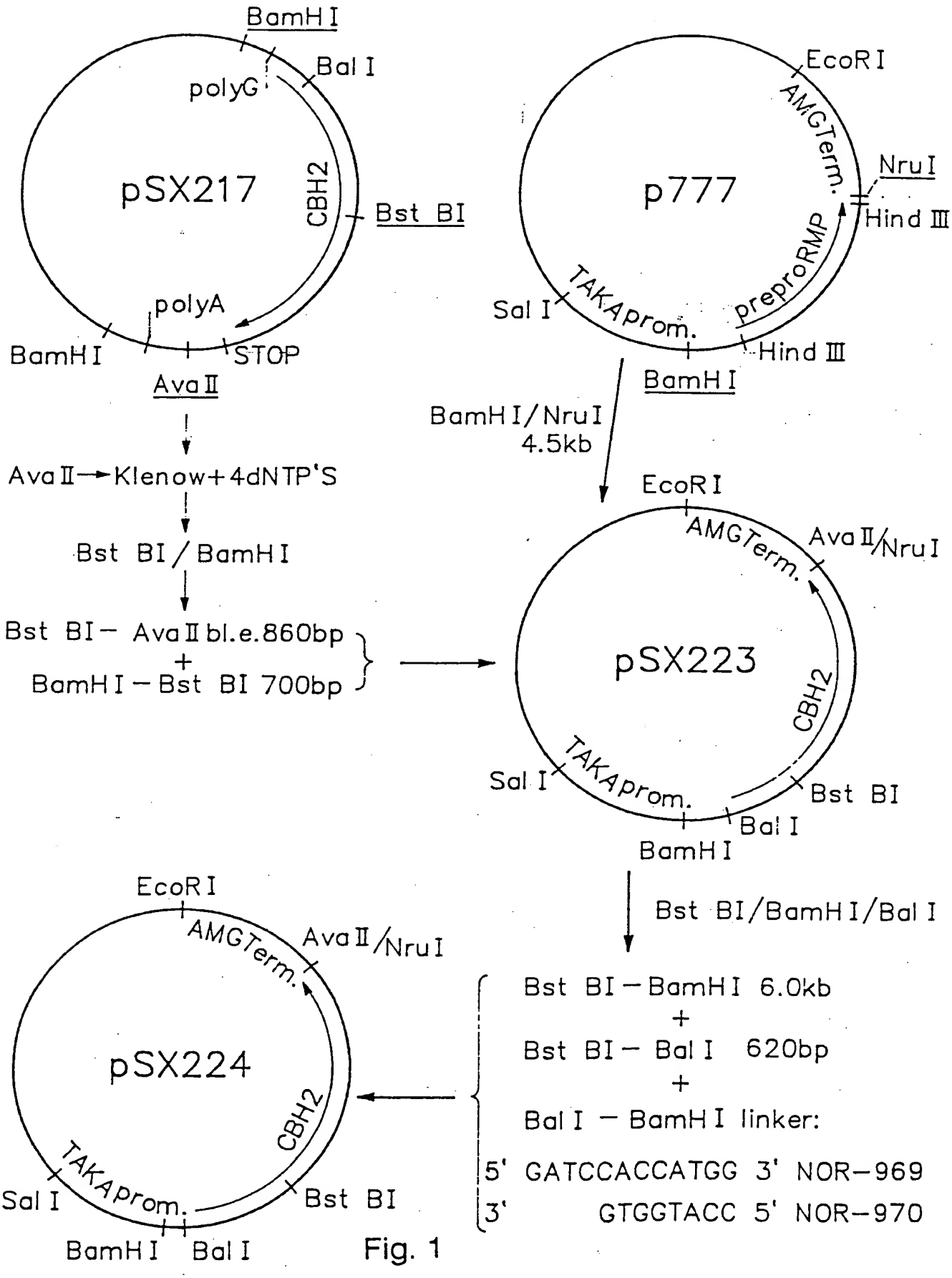
BamHI
BalI
polyG
pSX217
CBH2
Bst BI
polyA
BamHI
STOP
AvaII
AvaII→Klenow+4dNTP'S
Bst BI / BamHI
Bst BI – AvaII bl.e.860bp
+
BamHI – Bst BI 700bp
EcoRI
AMGTerm.
NruI
p777
Hind III
TAKAprom.
preproRMP
Sal I
Hind III
BamHI
BamHI/NruI
4.5kb
EcoRI
AMGTerm.
AvaII/NruI
pSX223
CBH2
TAKAprom.
Sal I
Bst BI
BalI
BamHI
Bst BI/BamHI/BalI
Bst BI – BamHI 6.0kb
+
Bst BI – BalI 620bp
+
BalI – BamHI linker:
5' GATCCACCATGG 3' NOR–969
3' GTGGTACC 5' NOR–970
EcoRI
AMGTerm.
AvaII/NruI
pSX224
CBH2
TAKAprom.
Sal I
Bst BI
BamHI
BalI

Fig. 1

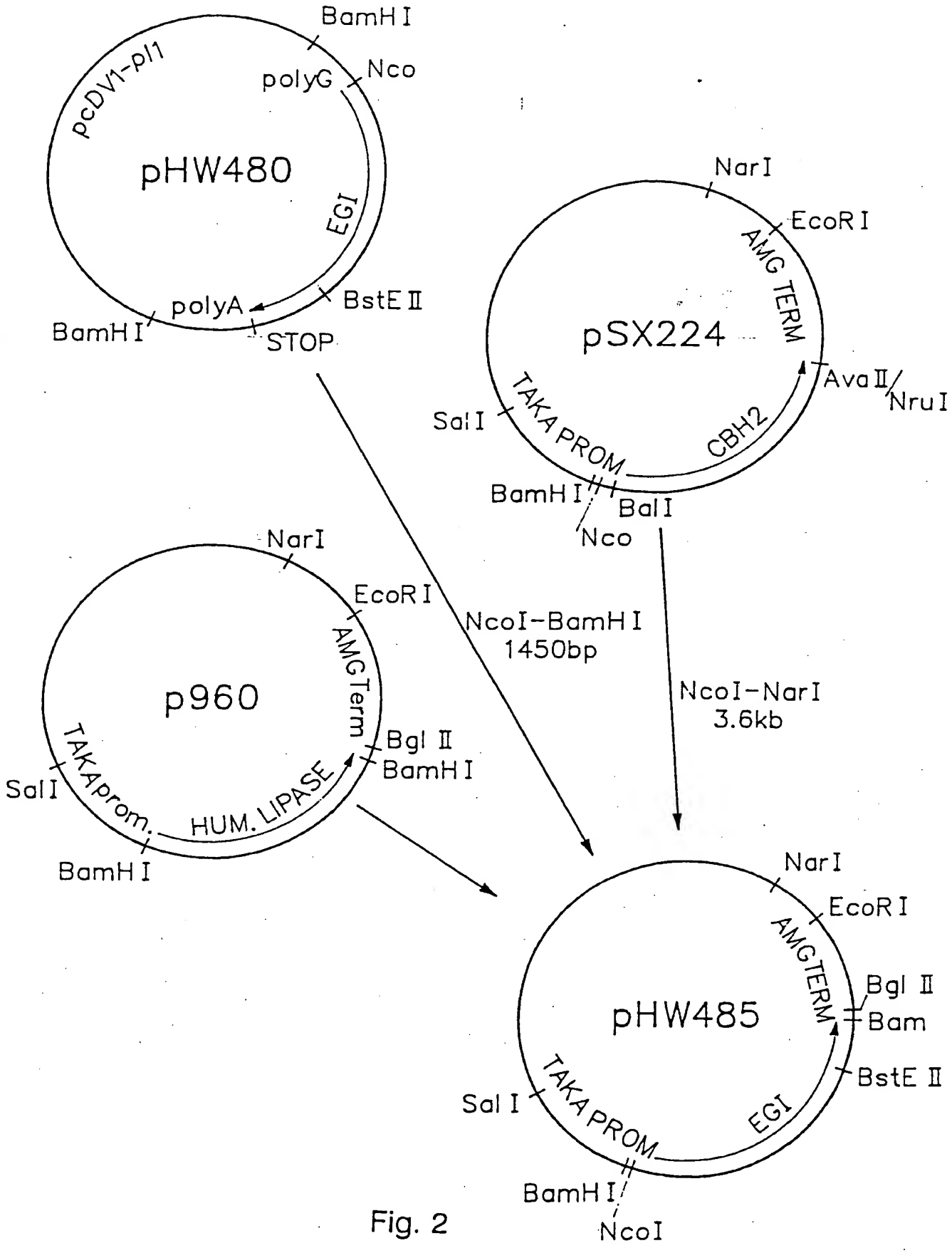
BamH I
pcDV1-pl1
polyG
Nco
pHW480
EGI
polyA
BstE II
BamH I
STOP
NarI
EcoR I
AMG TERM
pSX224
Ava II/
NruI
Sal I
TAKA PROM
CBH2
BamH I
Nco
Bal I
NarI
EcoR I
AMGTerm
p960
Bgl II
BamH I
Sal I
TAKAprom.
HUM. LIPASE
BamH I
NcoI-BamH I
1450bp
NcoI-NarI
3.6kb
NarI
EcoR I
AMGTERM
pHW485
Bgl II
Bam
BstE II
Sal I
TAKA PROM
EGI
BamH I
NcoI

Fig. 2

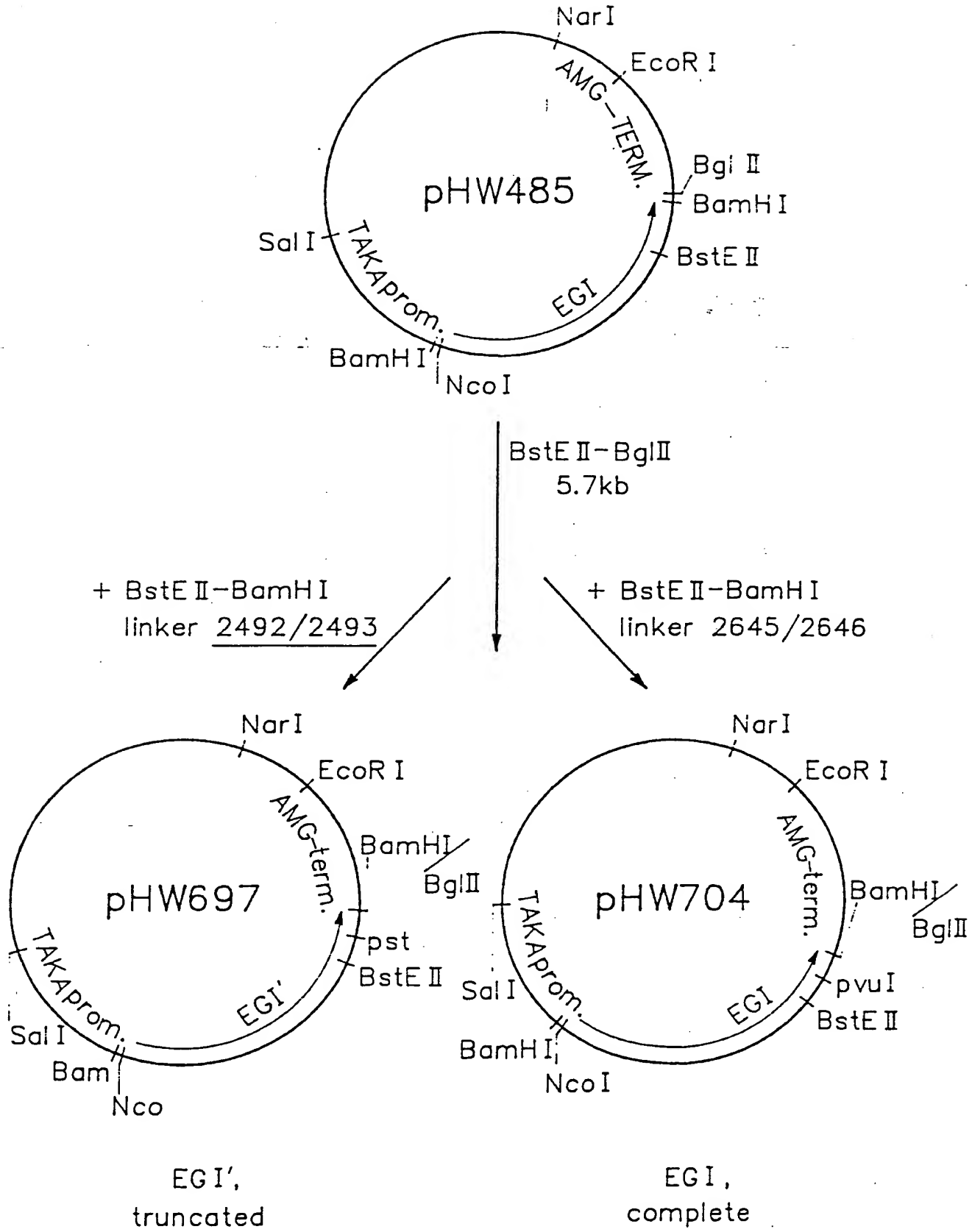
NarI
EcoR I
AMG-TERM.
Bgl II
BamH I
pHW485
BstE II
EGI
NcoI
BamH I
TAKAprom.
Sal I
BstE II-BglII
5.7kb
+ BstE II-BamH I
linker 2492/2493
+ BstE II-BamH I
linker 2645/2646
NarI
EcoR I
AMG-term.
BamHI
BglII
pHW697
pst
BstE II
EGI'
TAKAprom.
Sal I
Bam
Nco
NarI
EcoR I
AMG-term.
BamHI
BglII
pHW704
TAKAprom.
Sal I
BamHI
NcoI
EGI
pvuI
BstE II
EGI',
truncated
EGI,
complete

Fig. 3

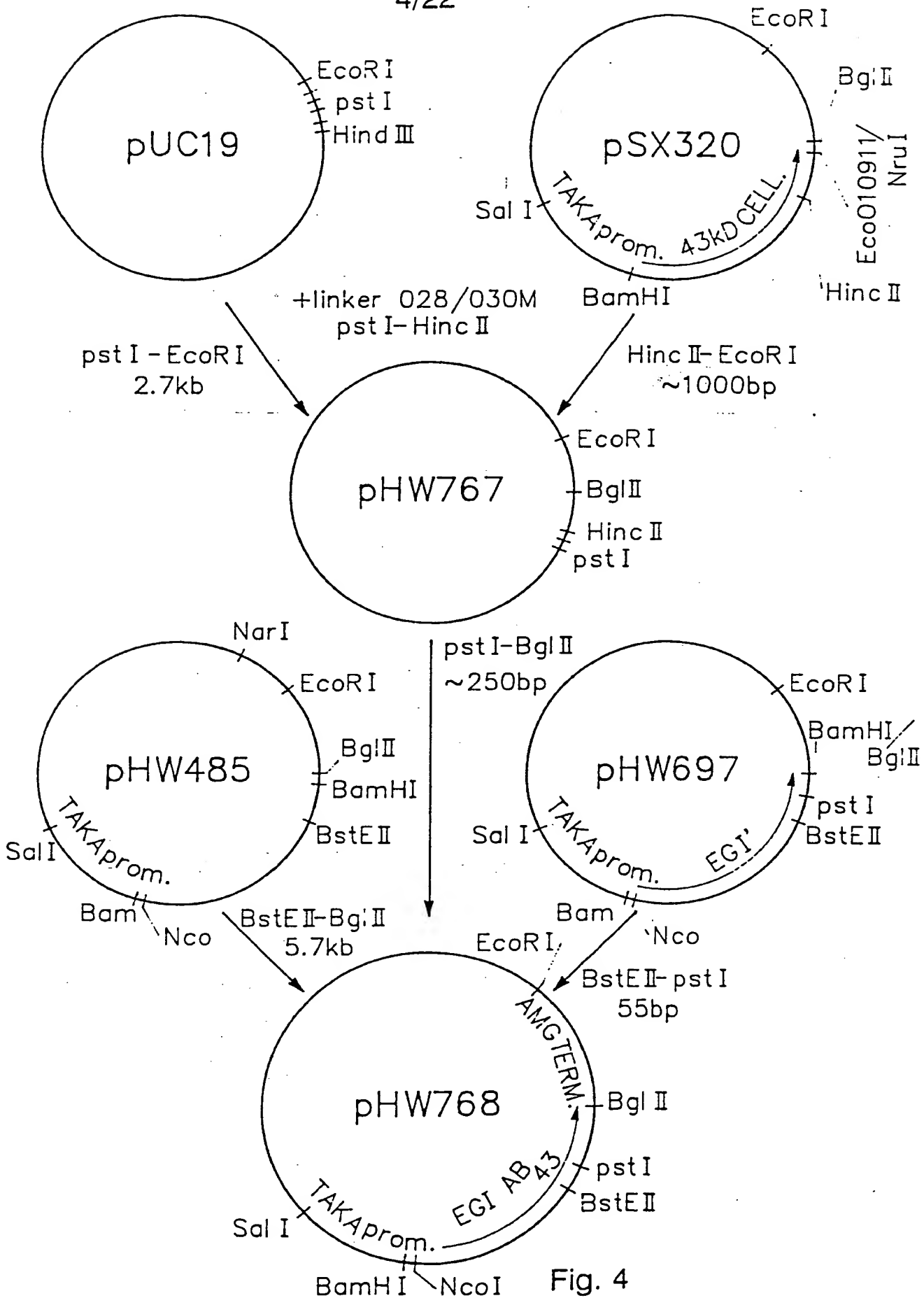
pUC19
EcoR I
pst I
Hind III
pSX320
EcoR I
Bgl II
Eco0109I/
NruI
Hinc II
BamHI
Sal I
TAKAprom.
43kD CELL.
+linker 028/030M
pst I-Hinc II
pst I - EcoR I
2.7kb
Hinc II- EcoR I
~1000bp
pHW767
EcoR I
BglII
Hinc II
pst I
pst I-Bgl II
~250bp
pHW485
NarI
EcoR I
BglII
BamHI
BstEII
Sal I
TAKAprom.
Bam
Nco
BstEII-BglII
5.7kb
pHW697
EcoR I
BamHI
BglII
pst I
BstEII
Sal I
TAKAprom.
EGI'
Bam
Nco
EcoR I
BstEII-pst I
55bp
pHW768
AMGTERM.
Bgl II
pst I
BstEII
EGI AB43
Sal I
TAKAprom.
BamH I
NcoI

Fig. 4

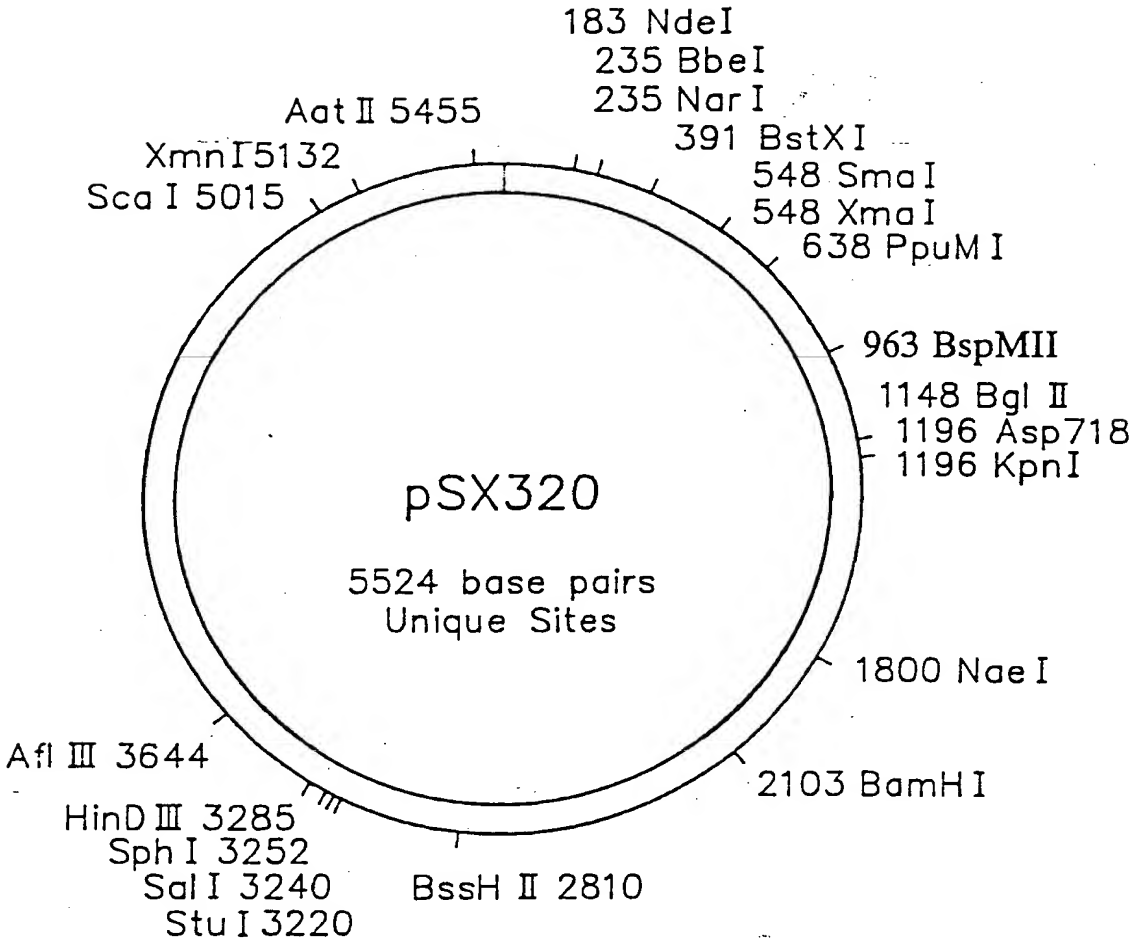
183 NdeI
235 BbeI
235 Nar I
391 BstX I
548 Sma I
548 Xma I
638 PpuM I
963 BspMII
1148 Bgl II
1196 Asp718
1196 KpnI
1800 Nae I
2103 BamH I
BssH II 2810
Stu I 3220
Sal I 3240
Sph I 3252
HinD III 3285
Afl III 3644
Sca I 5015
Xmn I 5132
Aat II 5455
pSX320
5524 base pairs
Unique Sites

Fig. 5

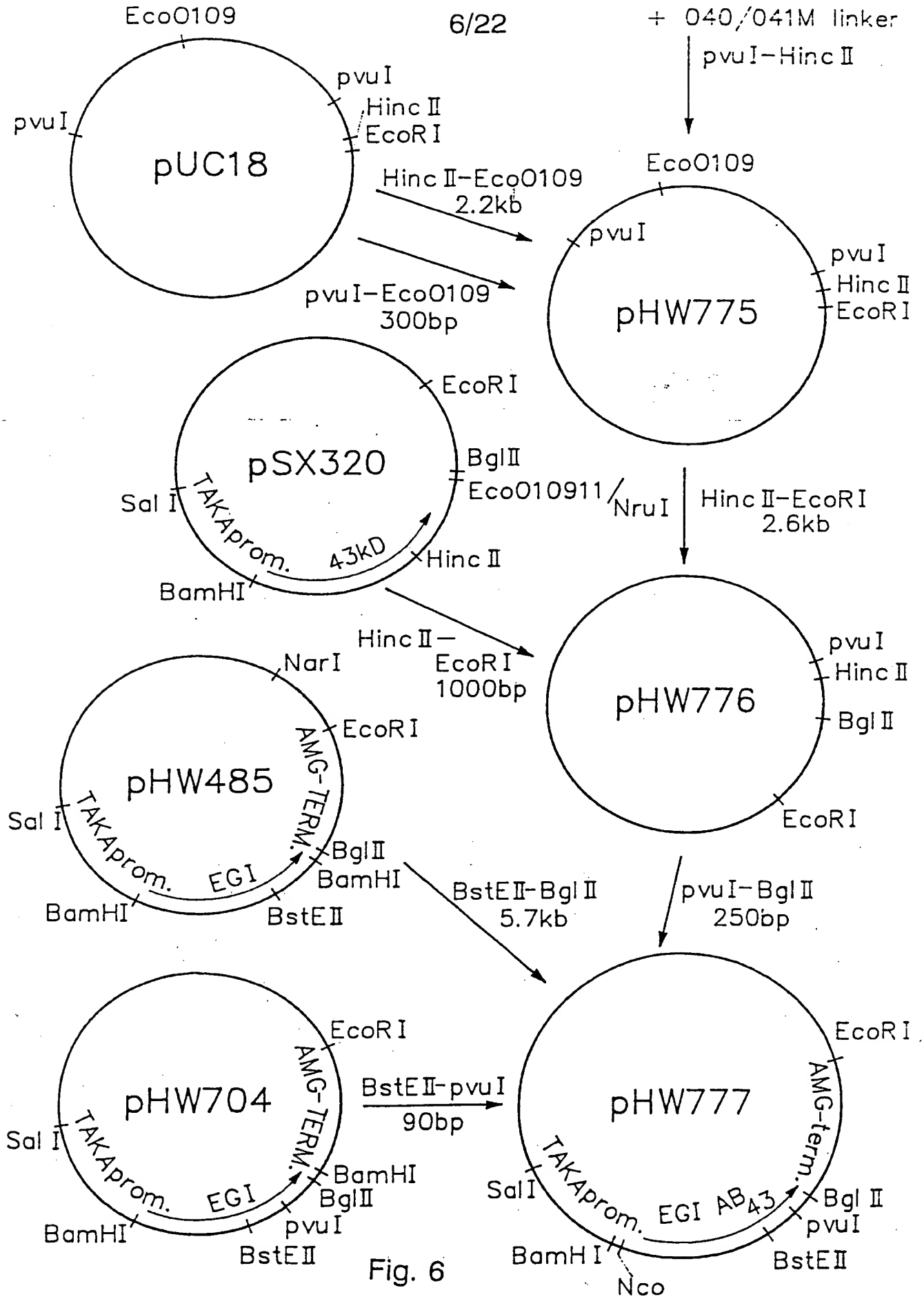
Eco0109
pvuI
pvuI
Hinc II
EcoR I
pUC18
+ 040/041M linker
pvuI-Hinc II
Hinc II-Eco0109
2.2kb
pvuI-Eco0109
300bp
Eco0109
pvuI
pvuI
Hinc II
EcoR I
pHW775
EcoR I
pSX320
Bgl II
Eco010911/NruI
Sal I
TAKAprom.
43kD
Hinc II
BamHI
Hinc II-EcoR I
2.6kb
Hinc II-
EcoR I
1000bp
pvuI
Hinc II
Bgl II
EcoR I
pHW776
NarI
EcoR I
AMG-TERM.
pHW485
Sal I
TAKAprom.
EGI
BamHI
BamHI
BstEII
BstEII-Bgl II
5.7kb
pvuI-Bgl II
250bp
EcoR I
AMG-TERM.
pHW704
BstEII-pvuI
90bp
Sal I
TAKAprom.
EGI
BamHI
BamHI
Bgl II
pvuI
BstEII
EcoR I
AMG-term.
pHW777
Sal I
TAKAprom.
EGI
AB43
Bgl II
pvuI
BamH I
Nco
BstEII

Fig. 6

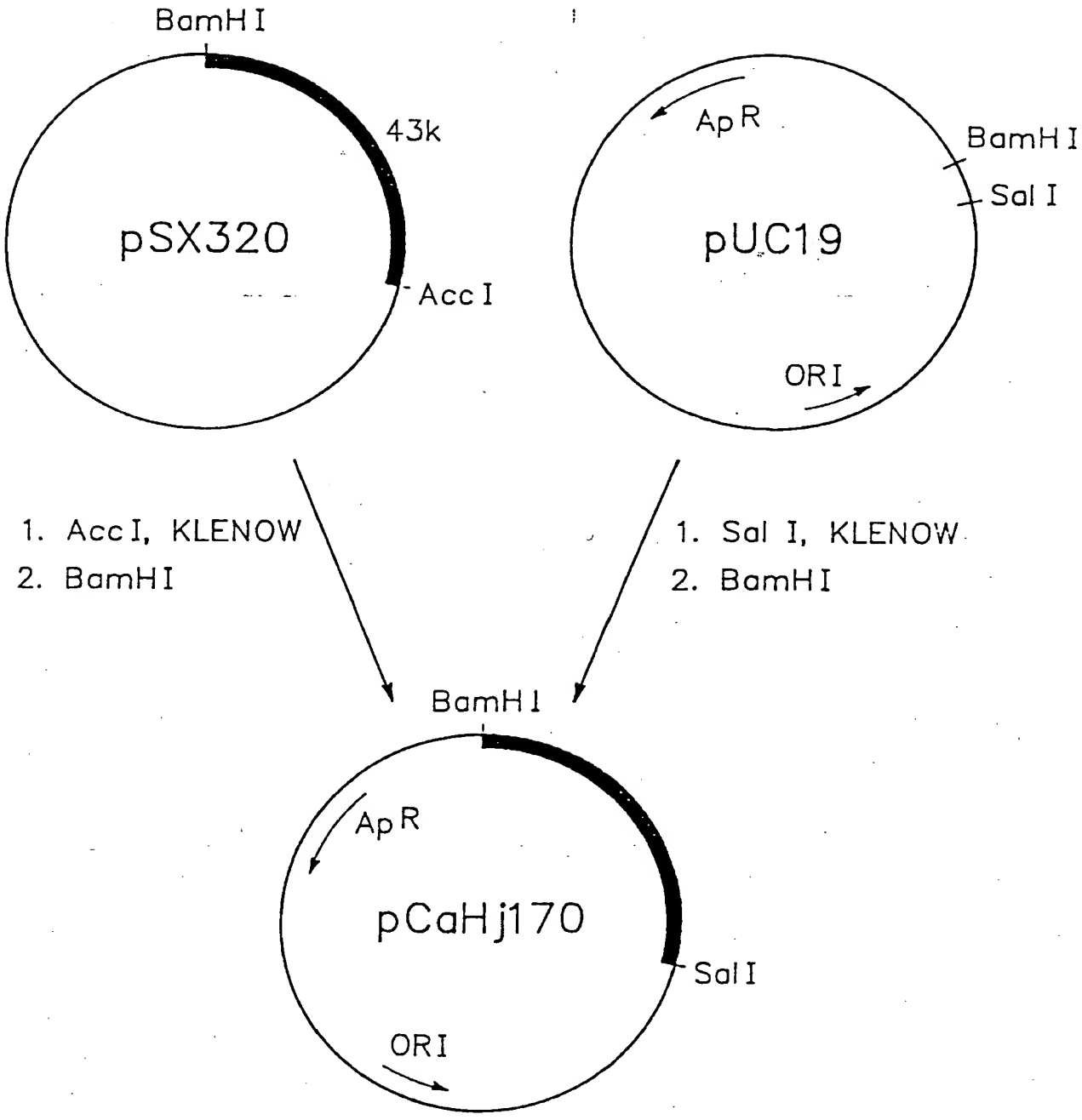
BamH I
43k
pSX320
Acc I
Ap R
BamH I
Sal I
pUC19
ORI
1. Acc I, KLENOW
2. BamHI
1. Sal I, KLENOW
2. BamHI
BamH I
Ap R
pCaHj170
Sal I
ORI

Fig. 7

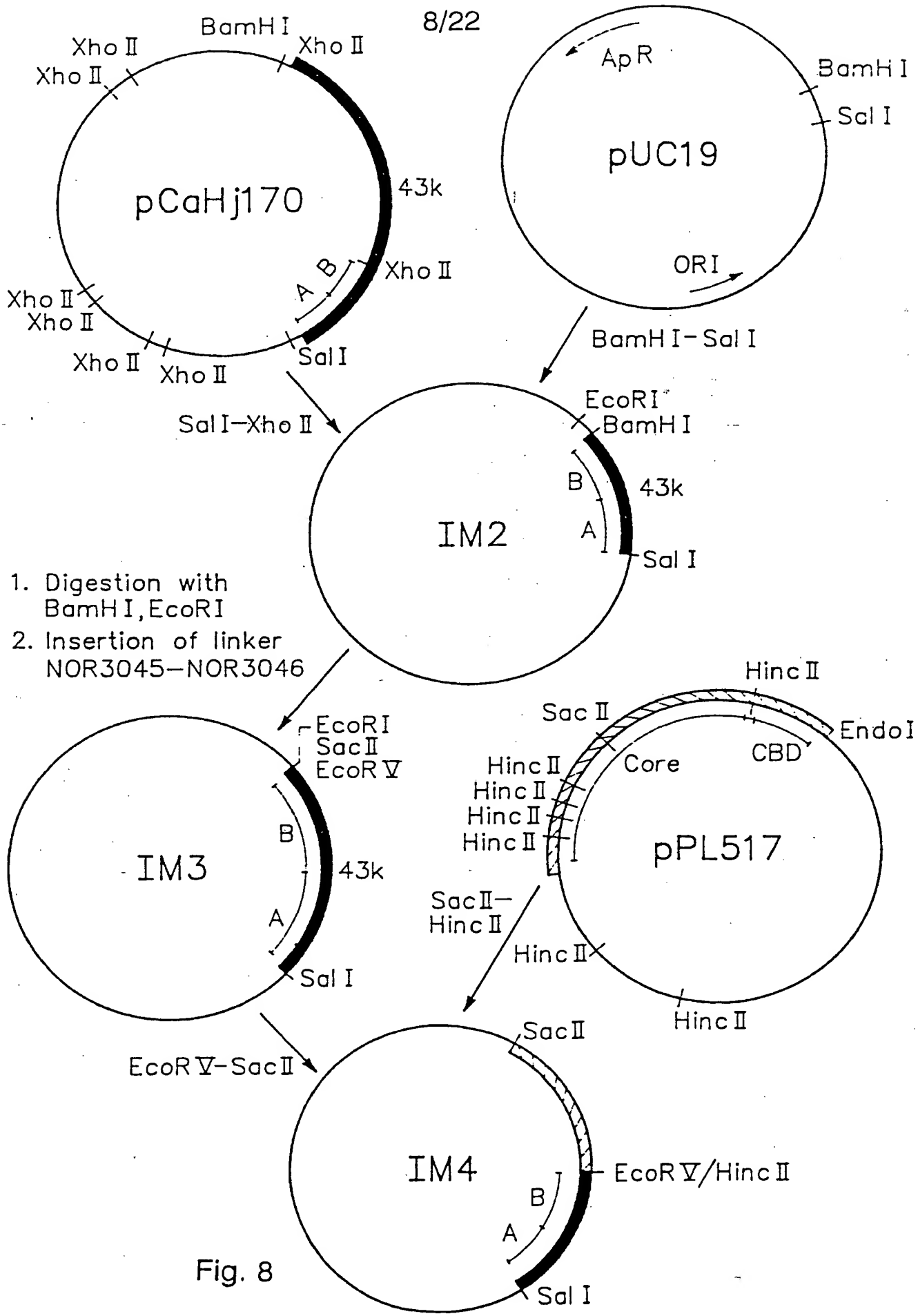
BamH I
Xho II
Xho II
Xho II
pCaHj170
43k
A
B
Xho II
Xho II
Xho II
Xho II
Sal I
ApR
BamH I
Sal I
pUC19
ORI
BamH I - Sal I
Sal I - Xho II
EcoRI
BamH I
B
43k
IM2
A
Sal I
1. Digestion with BamH I, EcoRI
2. Insertion of linker NOR3045-NOR3046
EcoRI
Sac II
EcoR V
B
IM3
43k
A
Sal I
Hinc II
Sac II
Endo I
CBD
Core
Hinc II
Hinc II
Hinc II
Hinc II
pPL517
Sac II -
Hinc II
Hinc II
Hinc II
Sac II
EcoR V - Sac II
IM4
EcoR V/Hinc II
B
A
Sal I

Fig. 8

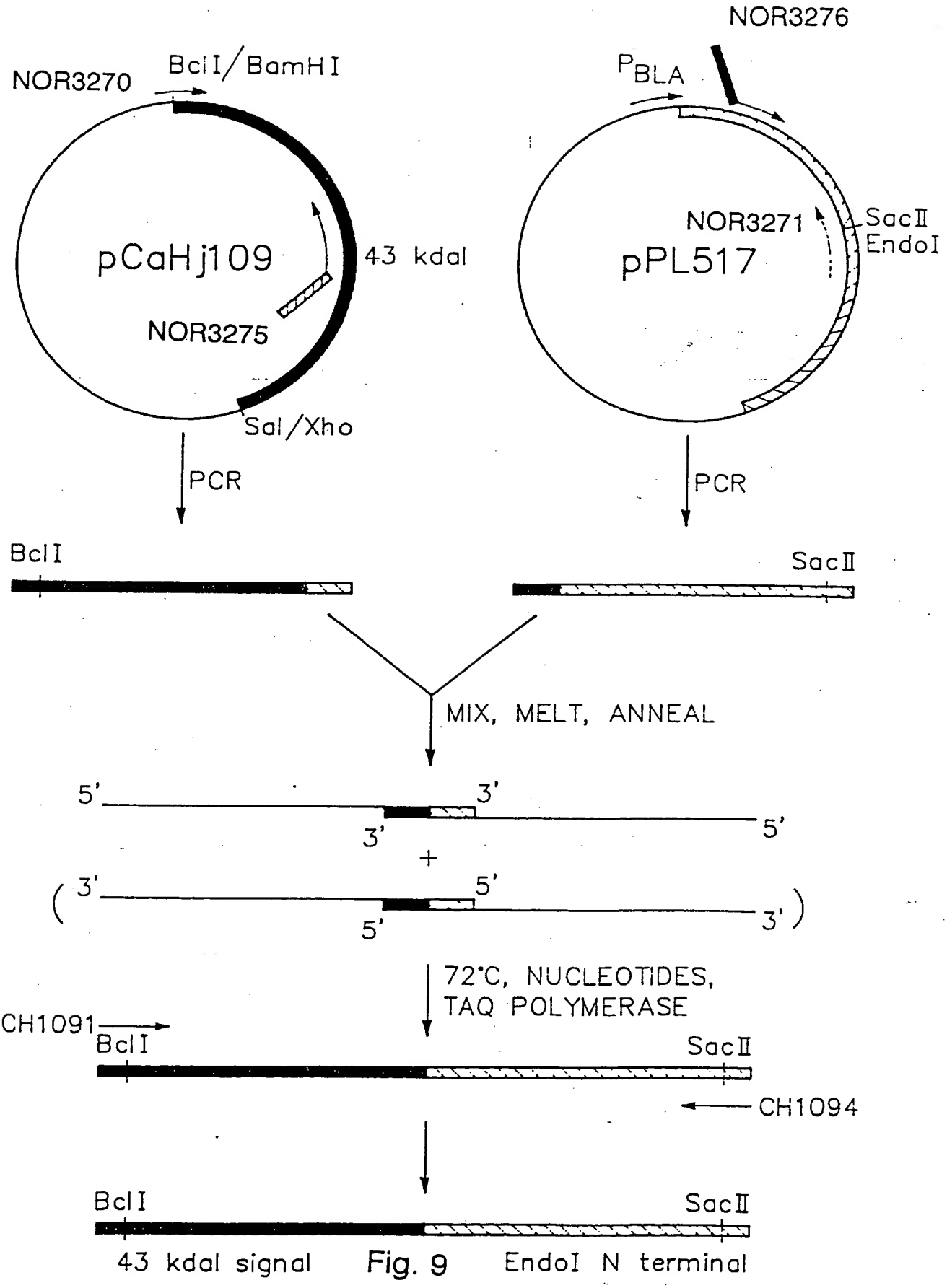
NOR3276
NOR3270
BclI/BamHI
P_{BLA}
pCaHj109
43 kdal
NOR3275
Sal/Xho
NOR3271
SacII
EndoI
pPL517
PCR
PCR
BclI
SacII
MIX, MELT, ANNEAL
5'
3'
3'
5'
+
3'
5'
5'
3'
72°C, NUCLEOTIDES,
TAQ POLYMERASE
CH1091
BclI
SacII
CH1094
BclI
SacII
43 kdal signal
EndoI N terminal

Fig. 9

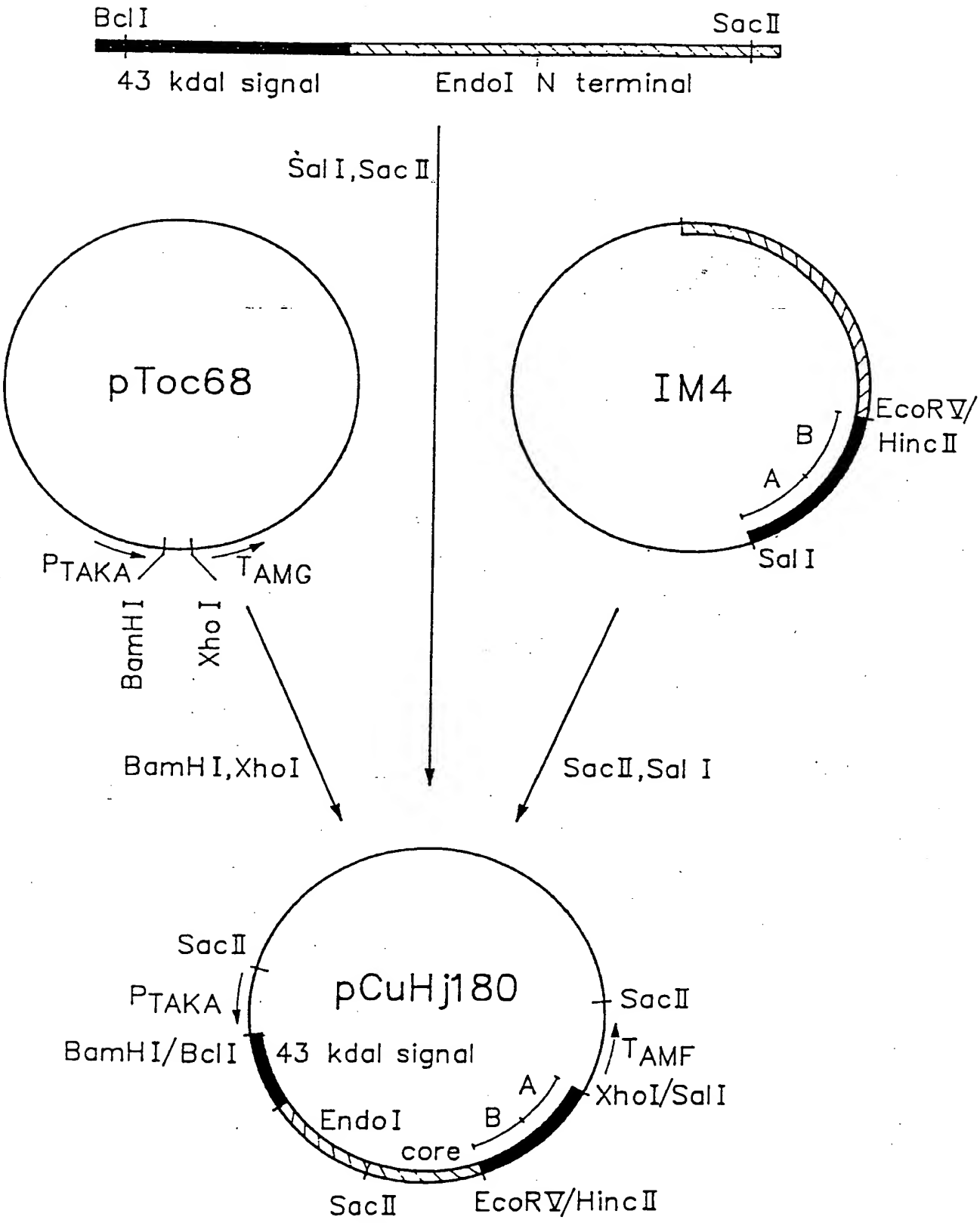
BclI
SacII
43 kdal signal
EndoI N terminal
SalI,SacII
pToc68
IM4
EcoRV/
HincII
B
A
SalI
PTAKA
TAMG
BamHI
XhoI
BamHI,XhoI
SacII,SalI
SacII
pCuHj180
SacII
PTAKA
BamHI/BclI
43 kdal signal
TAMF
XhoI/SalI
A
EndoI
B
core
SacII
EcoRV/HincII

Fig. 10

```
agaccggaattcgcggccgccatctatccaacggtctagcttcacttcacaatgtatcgc
                                                        M  Y  R
atcgtcgcaaccgcctcggctcttattgccgctgctcgggctcaacaggtctgctctttg
I  V  A  T  A  S  A  L  I  A  A  A  R  A  Q  Q  V  C  S  L
aacaccgagaccaagcctgccttgacctggtccaagtgtacatccagcggctgcagcgat
N  T  E  T  K  P  A  L  T  W  S  K  C  T  S  S  G  C  S  D
gtcaagggctccgttgttattgatgccaactggcgatggactcaccagacttctgggtct
V  K  G  S  V  V  I  D  A  N  W  R  W  T  H  Q  T  S  G  S
accaactgttacaccggaaacaagtgggacacctccatctgcactgatggcaagacctgc
T  N  C  Y  T  G  N  K  W  D  T  S  I  C  T  D  G  K  T  C
GCCGAAAAGTGCTGTCTTGATGGCGCCGACTATTCTGGTACCTACGGAATCACCTCCAGC
A  E  K  C  C  L  D  G  A  D  Y  S  G  T  Y  G  I  T  S  S
ggcaaccagctcagtcttggattcgtcaccaacggtccctacagcaagaacatcggcagc
G  N  Q  L  S  L  G  F  V  T  N  G  P  Y  S  K  N  I  G  S
cgaacctacctcatggagaacgagaacaccatccagatgttccagcttctgggcaacgag
R  T  Y  L  M  E  N  E  N  T  Y  Q  M  F  Q  L  L  G  N  E
ttcacctttgatgtcgatgtctctggtatcggctgcggtctgaacggtgcccctcacttc
F  T  F  D  V  D  V  S  G  I  G  C  G  L  N  G  A  P  H  F
gtcagcatggacgaggatggtggcaaggccaagtactccggaaacaaggccggagccaag
V  S  M  D  E  D  G  G  K  A  K  Y  S  G  N  K  A  G  A  K
tacggaactggcTACtGTGATGcccAgTGCCCTCGTGATGTCAAGTTCATCAACGGAGTT
Y  G  T  G  Y  C  D  A  Q  C  P  R  D  V  K  F  I  N  G  V
GCCAACTCTGAGGGCTGGAAGCCCTCTGACAGTGATGTCAACGCtggtgttggtaatctg
A  N  S  E  G  W  K  P  S  D  S  D  V  N  A  G  V  G  N  L
ggcacctgctgccccgagatggatatctgggaggccaactccatctccaccgccttcact
G  T  C  C  P  E  M  D  I  W  E  A  N  S  I  S  T  A  F  T
cctcatccttgcaccaagctcacacagcactcttgcactggcgactcttgtggtggaacc
P  H  P  C  T  K  L  T  Q  H  S  C  T  G  D  S  C  G  G  T
tactctagtgaccgatatggcggtacttgcgatgccgacggttgtgatttcaatgcctac
Y  S  S  D  R  Y  G  G  T  C  D  A  D  G  C  D  F  N  A  Y
cgtcagggcaacaagaccttctacggtcctggatccaacttcaacatcgacaccaccaag
R  Q  G  N  K  T  F  Y  G  P  G  S  N  F  N  I  D  T  T  K
aagatgactgttgtcactcagttccacaagggcAGCAAcGGACGTCTTTCTGAGATCACC
K  M  T  V  V  T  Q  F  H  K  G  S  N  G  R  L  S  E  I  T
CGTCTGTACGTCCAGAACGGCAAGGTCATTGCCAACTCAGAGTCCAAGATTGCAGGCAAC
R  L  Y  V  Q  N  G  K  V  I  A  N  S  E  S  K  I  A  G  N
CCCGGTAGCTCTCTCACCTCTGACTTCTGCTCCAAGCAGAAGAGCGTCTTTGGCGATATC
P  G  S  S  L  T  S  D  F  C  S  K  Q  K  S  V  F  G  D  I
GATGACTTCTCTAAGAAGGGTGGCTGGAACGgCATGAGCGATGCTCTCTCTGCCCCTATG
D  D  F  S  K  K  G  G  W  N  G  M  S  D  A  L  S  A  P  M
GTTCTTGTTATGTCTCTCTGGCACGACCACCACTCCAAcATGCTcTGGCTgGACtctacc
V  L  V  M  S  L  W  H  D  H  H  S  N  M  L  W  L  D  S  T
tacccaaccgactctaccaaggttggatctcaacgaggttcttgcgctaccacctctggc
Y  P  T  D  S  T  K  V  G  S  Q  R  G  S  C  A  T  T  S  G
aagccctccgaccttgagcgagatgttcccaactccaaggtttccttctccaacatcaAG
K  P  S  D  L  E  R  D  V  P  N  S  K  V  S  F  S  N  I  K
TTCGGTCCCATCGGAAGCACCTACAAGAGCGACGGCACCACCCCCAACCCCCCTgCCAGC
F  G  P  I  G  S  T  Y  K  S  D  G  T  T  P  N  P  P  A  S
AGCAGCACCACTGGTTCTTCCACTCCCACCAACCCCCCTGCCGGTAGCGTCGACCAATGG
S  S  T  T  G  S  S  T  P  T  N  P  P  A  G  S  V  D  Q  W
GGACAgTGcGGTGGCCAgaactacagcggccccacgacctgcaagtctcctttcacctgc
G  Q  C  G  G  Q  N  Y  S  G  P  T  T  C  K  S  P  F  T  C
aagaagatcaacgacttctactcccagtgtcagtaaaggggctgccgagctatctagcat
K  K  I  N  D  F  Y  S  Q  C  Q  .
gagattgagaaacgatgtgatgagtggacgatcaaggagaagtgtgtggatgatatgaac

ttgatgtgggaggac
```

Fig. 11

```
 gaattcgcggccgcctgcttcgaagcatcagctcattgagatcagtcaaaatgcatacc
                                                        M  H  T
ctttcggttctcctcgctctcgctcccgtgtccgcccttgctcaggctcccatctgggga
L  S  V  L  L  A  L  A  P  V  S  A  L  A  Q  A  P  I  W  G
cagtgcggtggcaatggttggaccggtgctacaacctgcgctagtggtctgaagtgtgag
Q  C  G  G  N  G  W  T  G  A  T  T  C  A  S  G  L  K  C  E
aagatcaacgactggtactatcagtgtgttcctggatctggaggatctgaaccccagcct
K  I  N  D  W  Y  Y  Q  C  V  P  G  S  G  G  S  E  P  Q  P
tcgtcaactcagggtggtggcactcctcagcctactggcggtaacagcggcggcactggt
S  S  T  Q  G  G  G  T  P  Q  P  T  G  G  N  S  G  G  T  G
ctcgacgccaaattcaaggccaagggcaagcagtactttggtaccgagattgaccactac
L  D  A  K  F  K  A  K  G  K  Q  Y  F  G  T  E  I  D  H  Y
caccttaacaacaatcctctgatcaacattgtcaaggcccagtttggccaagtgacatgc
H  L  N  N  N  P  L  I  N  I  V  K  A  Q  F  G  Q  V  T  C
gagaacagcatgaagtgggatgccattgagccttcacgcaactccttcaccttcagtaac
E  N  S  M  K  W  D  A  I  E  P  S  R  N  S  F  T  F  S  N
gctgacaaggtcgtcgacttcgccactcagaacggcaagctcatccgtggCCACACTCTT
A  D  K  V  V  D  F  A  T  Q  N  G  K  L  I  R  G  H  T  L
CTCTGGCACTCTCAGCTGCCTCAGTGGGTTCAGAACATCAACGATCGCTCTACCCTCACC
L  W  H  S  Q  L  P  Q  W  V  Q  N  I  N  D  R  S  T  L  T
GCGGTCATCGAGAACCACGTCAAGACCATGGTCACCCGCTACAAGGGCAAGATCCTCCAG
A  V  I  E  N  H  V  K  T  M  V  T  R  Y  K  G  K  I  L  Q
TGGGACGTTGTCAACAACGAGATCTTCGCTGAGGACGGTAACCTCCGCGACAGTGTCTTC
W  D  V  V  N  N  E  I  F  A  E  D  G  N  L  R  D  S  V  F
AGCCGAGTTCTCGGTGAGGACTTTGTCGGTATTGCTTTCCGCGCTGCCCGCGCCGCTGAT
S  R  V  L  G  E  D  F  V  G  I  A  F  R  A  A  R  A  A  D
CCCGCTGCCAAGCTCTACATCAACGATTATAACCTCGACAAGTCCGACTATGCTAAGGTC
P  A  A  K  L  Y  I  N  D  Y  N  L  D  K  S  D  Y  A  K  V
ACCCGCGGAATGGTCGCTCACGTTAATAAGTGGATTGCTGCTGGTATTCCCATCGACGGT
T  R  G  M  V  A  H  V  N  K  W  I  A  A  G  I  P  I  D  G
ATTGGATCTCAGGGCCATCTTGCTGCTCCTAGTGGCTGGAACCCTGCCTCTGGTGTTCCT
I  G  S  Q  G  H  L  A  A  P  S  G  W  N  P  A  S  G  V  P
GCTGCTCTCCGAGCTCTTGCCGCCTCGGACGCCAAGGAGATTGCTATcactgagcttgat
A  A  L  R  A  L  A  A  S  D  A  K  E  I  A  I  T  E  L  D
attgccggtgccagtgctaacgattaccttactgtcatgaacgcttgccttgccgttccc
I  A  G  A  S  A  N  D  Y  L  T  V  M  N  A  C  L  A  V  P
aagtgtgtcggcatcactgtctggggtgtctctgacaaggactcgtggcgacctggtgac
K  C  V  G  I  T  V  W  G  V  S  D  K  D  S  W  R  P  G  D
aacccccctcctctacgacagcaactaccagcccaaggctgctttcaatgccttggctaac
N  P  L  L  Y  D  S  N  Y  Q  P  K  A  A  F  N  A  L  A  N
gctctgtgagctgttgttgatgtatgtcgctggatcatacaacgaaacgtcctagttgga
A  L  .
taaagcgttgatggtagaatgat
```

Fig. 12

```
gaattcgcggccgcctagataagtcactacctgatctctgaataatctttcatcatgaag
                                                          M  K
tctctctcactcatcctctcagccctggctgtccaggtcgctgttgctcaaacccccgac
S  L  S  L  I  L  S  A  L  A  V  Q  V  A  V  A  Q  T  P  D
aaggccaaggagcagcaccccaagctcgagacctaccgctgcaccaaggcctctggctgc
K  A  K  E  Q  H  P  K  L  E  T  Y  R  C  T  K  A  S  G  C
aagaagcaaaccaactacatcgtcgccgacgcaggtattcacggcattcgcagaagcgcc
K  K  Q  T  N  Y  I  V  A  D  A  G  I  H  G  I  R  R  S  A
GGCTGCGGTGACTGGGGTCAAAAGCCCAACGCCACAGCCTGCCCCGATGAGGCATCCTGC
G  C  G  D  W  G  Q  K  P  N  A  T  A  C  P  D  E  A  S  C
GCTAAGAACTGTATCCTCAGTGGTATGGACTCAAACGCTTACAAGAACGCTGGTATCACT
A  K  N  C  I  L  S  G  M  D  S  N  A  Y  K  N  A  G  I  T
ACTTCTGGCAACAAGCTTCGTCTTCAGCAGCTTATCAACAACCAGCTTGTTTCTCCTCGG
T  S  G  N  K  L  R  L  Q  Q  L  I  N  N  Q  L  V  S  P  R
GTTTATCTGCTTGAGGAGAACAAGAAGAAGTATGAGATGCTTCAGCTCACTGGTACTGAA
V  Y  L  L  E  E  N  K  K  K  Y  E  M  L  H  L  T  G  T  E
TTCTCTTTCGACGTTGAGATGGAGAAGCTTCCTTGTGGTATGAATGGTGCTTTGTACCTT
F  S  F  D  V  E  M  E  K  L  P  C  G  M  N  G  A  L  Y  L
TCCGAGATGCCACAGGATGGTGGTAAGAGCACGAGCCGAAACAGCAAGGCTGGTGCCTAC
S  E  M  P  Q  D  G  G  K  S  T  S  R  N  S  K  A  G  A  Y
TATGGTGCTGGATACTGTGATGCTCAGTGCTACGTCactcctttcATCAACGGAGTTGGC
Y  G  A  G  Y  C  D  A  Q  C  Y  V  T  P  F  I  N  G  V  G
AACATCAAGGGACAGGGTGTCTGCTGTAACGAGCTCGACATCTGGGAGGCCAACTCCCGC
N  I  K  G  Q  G  V  C  C  N  E  L  D  I  W  E  A  N  S  R
GCAACTCACATTGCTCCTCACCCTTGCAGCAAGCCCGGCCTCTACGGCTGCACAGGCGAT
A  T  H  I  A  P  H  P  C  S  K  P  G  L  Y  G  C  T  G  D
GAGTGCGGCAGCTCCGGTTTCTGCGACAAGGCCGGCTGCGGCTGGAACCACAACCGCATC
E  C  G  S  S  G  I  C  D  K  A  G  C  G  W  N  H  N  R  I
AACGTGACCGACTTCTACGGccgcggCAAGCAGTACAAGGTCGACAGCACCCGCAAGTTC
N  V  T  D  F  Y  G  R  G  K  Q  Y  K  V  D  S  T  R  K  F
ACCGTGACATCTCAGTTCGTCGCCAACAAGCAGGGTGATCTCATCGAGCTGCACCGCCAC
T  V  T  S  Q  F  V  A  N  K  Q  G  D  L  I  E  L  H  R  H
TACATCCAGGACAACAAGGTCAtcgagtctgctgtcgtcaacatctccggccctcccaag
Y  I  Q  D  N  K  V  I  E  S  A  V  V  N  I  S  G  P  P  K
atcaacttcatcaatgacaagtactgcgctgccaccggcgccaacgagtacatgcgcctc
I  N  F  I  N  D  K  Y  C  A  A  T  G  A  N  E  Y  M  R  L
ggcggtactaagcaaatgggcgatgccatgtcccgcggaatggttctcgccatgagcgtc
G  G  T  K  Q  M  G  D  A  M  S  R  G  M  V  L  A  M  S  V
tggtggagcgagggtgatttcatggcctggttggatcagggtgttgctggaccctgtgac
W  W  S  E  G  D  F  M  A  W  L  D  Q  G  V  A  G  P  C  D
gccaccgagggcgatcccaagaacatcgtcaaggtgcagcccaaccctgaagtgacattt
A  T  E  G  D  P  K  N  I  V  K  V  Q  P  N  P  E  V  T  F
agcaacatcagaattggagagattggatctacttcatcggtcaaggctcctgcgtatcct
S  N  I  R  I  G  E  I  G  S  T  S  S  V  K  A  P  A  Y  P
ggtcctcaccgcttgtaaaaacatcaaacaacaccgtgtccaatatggATCTTAGTGTCC
G  P  H  R  L  .
ACTTGCTGGGAAGCTATTGGAGCACATATGCAAAACAGATGTCCACTAGCTTGACACGTA

TGTCGGGGCAAAAAAATCTCTTTCTAGGATAGGAGAACATATTGGGTGTTTGGACTTGTA

TATAAATGATACATTTTTCATATTATATTATTTTCAACATATTTTATTTCACGAAAAAAA

AAAAAAAAAAAAAAAAAAAAAAAA
```

Fig. 13

```
        10         20         30         40         50         60
        |          |          |          |          |          |
TTTCTTCGTCGAGCTCGAGTCGTCCGCCGTCTCCTCCTCCTCCTCCTTCCAGTCTTTGAG

        70         80         90        100        110        120
        |          |          |          |          |          |
TTCCTTCGACCTGCAGCGTCCTGAACAACTCGCTCTAGCTCAACAACCATGGCTCGCGGT
                                                  MetAlaArgGly

       130        140        150        160        170        180
        |          |          |          |          |          |
ACCGCTCTCCTCGGCCTGACCGCGCTCCTCCTGGGGCTGGTCAACGGCCAGAAGCCTGGT
ThrAlaLeuLeuGlyLeuThrAlaLeuLeuLeuGlyLeuValAsnGlyGlnLysProGly

       190        200        210        220        230        240
        |          |          |          |          |          |
GAGACCAAGGAGGTTCACCCCCAGCTCACGACCTTCCGCTGCACGAAGAGGGGTGGTTGC
GluThrLysGluValHisProGlnLeuThrThrPheArgCysThrLysArgGlyGlyCys

       250        260        270        280        290        300
        |          |          |          |          |          |
AAGCCGGCGACCAACTTCATCGTGCTTGACTCGCTGTCGCACCCCATCCACCGCGCTGAG
LysProAlaThrAsnPheIleValLeuAspSerLeuSerHisProIleHisArgAlaGlu
```

Fig. 14A

```
          310         320         330         340         350         360
           |           |           |           |           |           |
GGCCTGGGCCCTGGCGGCTGCGGCGACTGGGGCAACCCGCCGCCCAAGGACGTCTGCCCG
GlyLeuGlyProGlyGlyCysGlyAspTrpGlyAsnProProProLysAspValCysPro

          370         380         390         400         410         420
           |           |           |           |           |           |
GACGTCGAGTCGTGCGCCAAGAACTGCATCATGGAGGGCATCCCCGACTACAGCCAGTAC
AspValGluSerCysAlaLysAsnCysIleMetGluGlyIleProAspTyrSerGlnTyr

          430         440         450         460         470         480
           |           |           |           |           |           |
GGCGTCACCACCAACGGCACCAGCCTCCGCCTGCAGCACATCCTCCCCGACGGCCGCGTC
GlyValThrThrAsnGlyThrSerLeuArgLeuGlnHisIleLeuProAspGlyArgVal

          490         500         510         520         530         540
           |           |           |           |           |           |
CCGTCGCCGCGTGTCTACCTGCTCGACAAGACGAAGCGCCGCTATGAGATGCTCCACCTG
ProSerProArgValTyrLeuLeuAspLysThrLysArgArgTyrGluMetLeuHisLeu

          550         560         570         580         590         600
           |           |           |           |           |           |
ACCGGCTTCGAGTTCACCTTCGACGTCGACGCCACCAAGCTGCCCTGCGGCATGAACAGC
ThrGlyPheGluPheThrPheAspValAspAlaThrLysLeuProCysGlyMetAsnSer

          610         620         630         640         650         660
           |           |           |           |           |           |
GCTCTGTACCTGTCCGAGATGCACCCGACCGGTGCCAAGAGCAAGTACAACTCCGGCGGT
AlaLeuTyrLeuSerGluMetHisProThrGlyAlaLysSerLysTyrAsnSerGlyGly
```

Fig. 14B

```
      670        680        690        700        710        720
       |          |          |          |          |          |
GCCTACTACGGTACTGGCTACTGCGATGCTCAGTGCTTCGTGACGCCCTTCATCAACGGC
AlaTyrTyrGlyThrGlyTyrCysAspAlaGlnCysPheValThrProPheIleAsnGly

      730        740        750        760        770        780
       |          |          |          |          |          |
TTGGGCAACATCGAGGGCAAGGGCTCGTGCTGCAACGAGATGGATATCTGGGAGGTCAAC
LeuGlyAsnIleGluGlyLysGlySerCysCysAsnGluMetAspIleTrpGluValAsn

      790        800        810        820        830        840
       |          |          |          |          |          |
TCGCGCGCCTCGCACGTGGTTCCCCACACCTGCAACAAGAAGGGCCTGTACCTTTGCGAG
SerArgAlaSerHisValValProHisThrCysAsnLysLysGlyLeuTyrLeuCysGlu

      850        860        870        880        890        900
       |          |          |          |          |          |
GGTGAGGAGTGCGCCTTCGAGGGTGTTTGCGACAAGAACGGCTGCGGCTGGAACAACTAC
GlyGluGluCysAlaPheGluGlyValCysAspLysAsnGlyCysGlyTrpAsnAsnTyr

      910        920        930        940        950        960
       |          |          |          |          |          |
CGCGTCAACGTGACTGACTACTACGGCCGGGGCGAGGAGTTCAAGGTCAACACCCTCAAG
ArgValAsnValThrAspTyrTyrGlyArgGlyGluGluPheLysValAsnThrLeuLys

      970        980        990        1000       1010       1020
       |          |          |          |          |          |
CCCTTCACCGTCGTCACTCAGTTCTTGGCCAACCGCAGGGGCAAGCTCGAGAAGATCCAC
ProPheThrValValThrGlnPheLeuAlaAsnArgArgGlyLysLeuGluLysIleHis
```

Fig. 14C

```
      1030       1040       1050       1060       1070       1080
        |          |          |          |          |          |
CGCTTCTACGTGCAGGACGGCAAGGTCATCGAGTCCTTCTACACCAACAAGGAGGGAGTC
ArgPheTyrValGlnAspGlyLysValIleGluSerPheTyrThrAsnLysGluGlyVal

      1090       1100       1110       1120       1130       1140
        |          |          |          |          |          |
CCTTACACCAACATGATCGATGACGAGTTCTGCGAGGCCACCGGCTCCCGCAAGTACATG
ProTyrThrAsnMetIleAspAspGluPheCysGluAlaThrGlySerArgLysTyrMet

      1150       1160       1170       1180       1190       1200
        |          |          |          |          |          |
GAGCTCGGCGCCACCCAGGGCATGGGCGAGGCCCTCACCCGCGGCATGGTCCTGGCCATG
GluLeuGlyAlaThrGlnGlyMetGlyGluAlaLeuThrArgGlyMetValLeuAlaMet

      1210       1220       1230       1240       1250       1260
        |          |          |          |          |          |
AGCATCTGGTGGGACCAGGGCGGCAACATGGAGTGGCTCGACCACGGCGAGGCCGGCCCC
SerIleTrpTrpAspGlnGlyGlyAsnMetGluTrpLeuAspHisGlyGluAlaGlyPro

      1270       1280       1290       1300       1310       1320
        |          |          |          |          |          |
TGCGCCAAGGGCGAGGGCGCCCCGTCCAACATTGTCCAGGTTGAGCCCTTCCCCGAGGTC
CysAlaLysGlyGluGlyAlaProSerAsnIleValGlnValGluProPheProGluVal

      1330       1340       1350       1360       1370       1380
        |          |          |          |          |          |
ACCTACACCAACCTCCGCTGGGGCGAGATCGGCTCGACCTACCAGGAGGTTCAGAAGCCT
ThrTyrThrAsnLeuArgTrpGlyGluIleGlySerThrTyrGlnGluValGlnLysPro
```

Fig. 14D

```
      1390       1400       1410       1420       1430       1440
        |          |          |          |          |          |
AAGCCCAAGCCCGGCCACGGCCCCCGGAGCGACTAAGTGGTGATGGGATAGAGGGATAGA
LysProLysProGlyHisGlyProArgSerAspEND

      1450       1460       1470       1480       1490       1500
        |          |          |          |          |          |
ATAGTGGATAGCACATAGATCGGCGGTTTTGGATAGTTTAATACATTCCGTTGCCGTTGT

      1510
        |
GAAAAAAAAA - poly-A
```

Fig. 14E

```
        10         20         30         40         50         60
        |          |          |          |          |          |
ATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGGCCCTT

METArgSerSerProLeuLeuProSerAlaValValAlaAlaLeuProValLeuAlaLeu
                43 kdal signalpeptide and N terminal

        70         80         90        100        110        120
        |          |          |          |          |          |
GCCGCTGATGGCAGGAGTGATGTCACTTTCACGATTAATACGCAGTCGGAACGTGCAGCG

AlaAlaAspGlyArgSerAspValThrPheThrIleAsnThrGlnSerGluArgAlaAla
   N terminal

       130        140        150        160        170        180
        |          |          |          |          |          |
ATCAGCCCCAATATTTACGGAACCAATCAGGATCTGAGCGGGACGGAGAACTGGTCATCC

IleSerProAsnIleTyrGlyThrAsnGlnAspLeuSerGlyThrGluAsnTrpSerSer

       190        200        210        220        230        240
        |          |          |          |          |          |
CGCAGGCTCGGAGGCAACCGGCTGACGGGTTACAACTGGGAGAACAACGCATCCAGCGCC

ArgArgLeuGlyGlyAsnArgLeuThrGlyTyrAsnTrpGluAsnAsnAlaSerSerAla

       250        260        270        280        290        300
        |          |          |          |          |          |
GGAAGGGACTGGCTTCATTACAGCGATGATTTTCTCTGCGGCAACGGTGGTGTTCCAGAC

GlyArgAspTrpLeuHisTyrSerAspAspPheLeuCysGlyAsnGlyGlyValProAsp
                    Endo 1 core

       310        320        330        340        350        360
        |          |          |          |          |          |
ACCGACTGCGACAAGCCGGGGGCGGTTGTTACCGCTTTTCACGATAAATCTTTGGAGAAT

ThrAspCysAspLysProGlyAlaValValThrAlaPheHisAspLysSerLeuGluAsn

       370        380        390        400        410        420
        |          |          |          |          |          |
GGAGCTTACTCCATTGTAACGCTGCAAATGGCGGGTTATGTGTCCCGGGATAAGAACGGT

GlyAlaTyrSerIleValThrLeuGlnMETAlaGlyTyrValSerArgAspLysAsnGly

       430        440        450        460        470        480
        |          |          |          |          |          |
CCAGTTGACGAGAGTGAGACGGCTCCGTCACCGCGTTGGGATAAGGTCGAGTTTGCCAAA

ProValAspGluSerGluThrAlaProSerProArgTrpAspLysValGluPheAlaLys
```

Fig. 15A

```
        490         500         510         520         530         540
AATGCGCCGTTCTCCCTTCAGCCTGATCTGAACGACGGACAAGTGTATATGGATGAAGAA
AsnAlaProPheSerLeuGlnProAspLeuAsnAspGlyGlnValTyrMETAspGluGlu

        550         560         570         580         590         600
GTTAACTTCCTGGTCAACCGGTATGGAAACGCTTCAACGTCAACGGGCATCAAAGCGTAT
ValAsnPheLeuValAsnArgTyrGlyAsnAlaSerThrSerThrGlyIleLysAlaTyr

        610         620         630         640         650         660
TCGCTGGATAACGAGCCGGCGCTGTGGTCTGAGACGCATCCAAGGATTCATCCGGAGCAG
SerLeuAspAsnGluProAlaLeuTrpSerGluThrHisProArgIleHisProGluGln

        670         680         690         700         710         720
TTACAAGCGGCAGAACTCGTCGCTAAGAGCATCGACTTGTCAAAGGCGGTGAAGAACGTC
LeuGlnAlaAlaGluLeuValAlaLysSerIleAspLeuSerLysAlaValLysAsnVal

        730         740         750         760         770         780
GATCCGCATGCCGAAATATTCGGTCCTGCCCTTTACGGTTTCGGCGCATATTTGTCTCTG
AspProHisAlaGluIlePheGlyProAlaLeuTyrGlyPheGlyAlaTyrLeuSerLeu

        790         800         810         820         830         840
CAGGACGCACCGGATTGGCCGAGTTTGCAAGGCAACTACAGCTGGTTTATCGATTACTAT
GlnAspAlaProAspTrpProSerLeuGlnGlyAsnTyrSerTrpPheIleAspTyrTyr

        850         860         870         880         890         900
CTGGATCAGATGAAGAATGCTCATACGCAGAACGGCAAAAGATTGCTCGATGTGCTGGAC
LeuAspGlnMETLysAsnAlaHisThrGlnAsnGlyLysArgLeuLeuAspValLeuAsp

        910         920         930         940         950         960
GTCCACTGGTATCCGGAAGCACAGGGCGGAGGCCAGCGAATCGTCTTTGGCGGGGCGGGC
ValHisTrpTyrProGluAlaGlnGlyGlyGlyGlnArgIleValPheGlyGlyAlaGly
```

Fig. 15B

970 980 990 1000 1010 1020

AATATCGATACGCAGAAGGCTCGCGTACAAGCGCCAAGATCGCTATGGGATCCGGCTTAC

AsnIleAspThrGlnLysAlaArgValGlnAlaProArgSerLeuTrpAspProAlaTyr

1030 1040 1050 1060 1070 1080

CAGGAAGACAGCTGGATCGGCACATGGTTTTCAAGCTACTTGCCCTTAATTCCGAAGCTG

GlnGluAspSerTrpIleGlyThrTrpPheSerSerTyrLeuProLeuIleProLysLeu

1090 1100 1110 1120 1130 1140

CAATCTTCGATTCAGACGTATTATCCGGGTACGAAGCTGGCGATCACAGAGTTCAGCTAC

GlnSerSerIleGlnThrTyrTyrProGlyThrLysLeuAlaIleThrGluPheSerTyr

1150 1160 1170 1180 1190 1200

GGCGGAGACAATCACATTTCGGGAGGCATAGCTACCGCGGACGCGCTCGGCATTTTTGGA

GlyGlyAspAsnHisIleSerGlyGlyIleAlaThrAlaAspAlaLeuGlyIlePheGly

1210 1220 1230 1240 1250 1260

AAATATGGCGTTTATGCCGCGAATTACTGGCAGACGGAGGACAATACCGATTATACCAGC

LysTyrGlyValTyrAlaAlaAsnTyrTrpGlnThrGluAspAsnThrAspTyrThrSer

1270 1280 1290 1300 1310 1320

GCTGCTTACAAGCTGTATCGCAACTACGACGGCAATAAATCGGGGTTCGGCTCGATCAAA

AlaAlaTyrLysLeuTyrArgAsnTyrAspGlyAsnLysSerGlyPheGlySerIleLys

1330 1340 1350 1360 1370 1380

GTGGACGCCGCTACGTCCGATACGGAGAACAGCTCGGTATACGCTTCGGTAACTGACGAG

ValAspAlaAlaThrSerAspThrGluAsnSerSerValTyrAlaSerValThrAspGlu

1390 1400 1410 1420 1430 1440

GAGAATTCCGAACTCCACCTGATCGTGCTGAATAAAAATTTCGACGATCCGATCAACGCT

GluAsnSerGluLeuHisLeuIleValLeuAsnLysAsnPheAspAspProIleAsnAla

Fig. 15C

```
      1450       1460       1470       1480       1490       1500
ACTTTCCAGCTGTCTGGTGATAAAACCTACACATCCGGGAGAGTATGGGGCTTCGACCAA
ThrPheGlnLeuSerGlyAspLysThrTyrThrSerGlyArgValTrpGlyPheAspGln

      1510       1520       1530       1540       1550       1560
ACCGGATCCGACATTACGGAACAAGCAGCTATAACGAATATTAACAACAATCAATTCACG
ThrGlySerAspIleThrGluGlnAlaAlaIleThrAsnIleAsnAsnAsnGlnPheThr

      1570       1580       1590       1600       1610       1620
TATACGCTTCCTCCATTGTCGGCTTACCACATTGTTCTGAAAGCGGATAGCACCGAACCG
TyrThrLeuProProLeuSerAlaTyrHisIleValLeuLysAlaAspSerThrGluPro

      1630       1640       1650       1660       1670       1680
GTCATCTCCGAGATCCCCTCCAGCAGCACCAGCTCTCCGGTCAACCAGCCTACCAGCACC
ValIleSerGluIleProSerSerSerThrSerSerProValAsnGlnProThrSerThr
   Linker                              43 kdal B region

      1690       1700       1710       1720       1730       1740
AGCACCACGTCCACCTCCACCACCTCGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGC
SerThrThrSerThrSerThrThrSerSerProProValGlnProThrThrProSerGly

      1750       1760       1770       1780       1790       1800
TGCACTGCTGAGAGGTGGGCTCAGTGCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGC
CysThrAlaGluArgTrpAlaGlnCysGlyGlyAsnGlyTrpSerGlyCysThrThrCys
                          43 kdal A region

      1810       1820       1830       1840       1850
GTCGCTGGCAGCACTTGCACGAAGATTAATGACTGGTACCATCAGTGCCTGTAG
ValAlaGlySerThrCysThrLysIleAsnAspTrpTyrHisGlnCysLeu---
```

Fig. 15D

INTERNATIONAL SEARCH REPORT

International Application No PCT/DK 91/00124

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC5: C 12 N 9/42, C 12 N 1/14

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
IPC5	C 12 N

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in Fields Searched[8]

SE,DK,FI,NO classes as above

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category *	Citation of Document,[11] with indication, where appropriate, of the relevant passages[12]	Relevant to Claim No.[13]
X	GENE, Vol. 51, 1987 Tuula T. Teeri: "Homologous domains in Trichoderma reesei cellulolytic enzymes: gene sequence and expression of cellobiohydrolase II", see page 43 - page 52 especially page 46 --	1-23
A	US, A, 4435307 (BARBESGAARD ET AL) 6 March 1984, see the whole document --	1-23
A	Chemical Abstracts, volume 105, no. 13, 29 September 1986, (Columbus, Ohio, US), Rao, Mala et al: "Purification characterization, and synergistic action of endoglucanases from Fusarium lini ", see page 307, abstract 110894p, & Biotechnol.Bioeng. 1986, 28(7), 1100-1105 --	1-23

* Special categories of cited documents:[10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance, the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance, the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
20th August 1991	1991 -08- 29
International Searching Authority	**Signature of Authorized Officer**
SWEDISH PATENT OFFICE	Yvonne Siösteen Yvonne Siösteen

Form PCT/ISA/210 (second sheet) (January 1985)

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
A	Chemical Abstracts, volume 109, no. 23, 5 December 1988, (Columbus, Ohio, US), Hayashida, Shinsaku et al: "Cellulases of Humicola insolens and Humicola grisea ", see page 295, abstract 207112c, & Methods Enzymol. 1988, 160(), 323-332 --	1-23
A	Chemical Abstracts, volume 105, no. 3, 21 July 1986, (Columbus, Ohio, US), Hayashida, Shinsaku et al: "Production and characteristics of Avicel-disintegrating endoglucanase from a protease-negative Humicola grisea var, thermoidea mutant ", see page 316, abstract 20820g, & Appl.Environ.Microbiol. 1986, 51(5), 1041-1046 --	1-23
A	Chemical Abstracts, volume 103, no. 7, 19 August 1985, (Columbus, Ohio, US), Rodionova, N.A. et al: "Characterization of endo-1,4-beta-glucanases of Geotrichum candidum ", see page 237, abstract 50208q, & Prikl.Biokhim.Mikrobiol. 1985, 21(3), 309- 317 --	1-23

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO.PCT/DK 91/00124

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the Swedish Patent Office EDP file on 91-06-27 The Swedish Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
US-A- 4435307	84-03-06	AT-B-	384442	87-11-10
		BE-A-	888632	81-10-29
		CH-A-	663511	87-12-31
		DE-A-	3117250	82-04-01
		FR-A-B-	2481712	81-11-06
		GB-A-B-	2075028	81-11-11
		JP-C-	1443072	88-06-08
		JP-A-	57023699	82-02-06
		JP-B-	61016316	86-04-30
		NL-A-	8102123	81-11-16